Partial Naivety in Present Bias: A Meta-analysis



Le Wu

University College

University of Oxford

A thesis submitted for the degree of

Master of Letters

Trinity 2021

Acknowledgements

Writing this thesis has been the greatest challenge in my life. While this is not the perfection I wanted to achieve, I tried my best with constant struggle and suffering.

My greatest gratitude goes to both of my supervisors, Dr. Abi Adams-Prassl and Dr. Severine Toussaert. They have done everything they can as supervisors and they are the sole reason I'm submitting this thesis right now. I can't emphasize enough how much they helped me in this process, both academically and personally.

I'd also like to show my appreciation to the University College and the department of Economics, especially to Caroline and Sanjay, for helping and guiding me through the most difficult period of my life.

Special mention to Yiliang Li and Yifan Zhang for being the best friends and housemates I can ever imagine.

Abstract

In this thesis, 47 estimates from 10 papers studying partial naivety in present bias were examined, focusing on elicitation and estimation of present bias, perceived present bias, and naivety. For these three parameters concerned, meta-analyses were conducted to determine the synthesis between estimates and potential sources of heterogeneity. Both time consistency and sophistication can be rejected from the synthesis result, while the heterogeneity analysis points out several possibilities causing the estimates' variation.

1 Introduction

In many scenarios, people are asked to make their choice given a set of benefits and costs at different times. The choice can be anything from a health-related decision (alcohol/smoking cessation, gym attendance) to a person's financial plans (credit card usage). This type of question falls under the intertemporal utility theory and has been examined for a long time. A growing number of empirical, especially experimental, studies suggested that people tend to show present bias in this context. Present bias is about the preference for immediate gratification, meaning that people who choose a bundle with immediate benefit and future cost may reverse their choice if the same bundle is moved from today to some periods later (and vice versa). This type of preference reversal is also denoted as time inconsistency.

For a long time, economists utilised the exponentially discounted model (EDM) (Samuelson, 1937) as a benchmark to study the inter-temporal choice problem. The discovery of present bias consists of a direct contradiction to the EDM model, and as a result, economists started to search for potential remedies. Within the set of newly proposed models, the quasi-hyperbolic discounting model (QHD) (Laibson, 1997) is the most frequently used one and is often used as an example of the PEEM (portable extension of existing models) principle proposed by Rabin, 2013. Compared with the EDM, the QHD model introduces one more discounting factor $\beta \in [0, 1]$, which measures the agent's preference between the present and future. The QHD is also denoted as the $\beta - \delta$ model due to its dual discounting factor structure.

The sole introduction of the QHD model only partially solves economists'

puzzles. To further explain this set of inconsistent choices, the idea of sophistication and naivety (ODR) (O'Donoghue & Rabin, 1999) was introduced. A person is sophisticated about his present bias if her belief of present bias $\hat{\beta}$ is equal to the value of her β, and she is a naif if she has a present bias but holds the incorrect belief that she is a time-consistent person.

Between these two extreme cases, in which people are either sophisticated or utterly naive about their present bias, it is reasonable to suspect that most people are partially aware of their present bias; they are denoted as the partial naivete or equally the partial sophisticate. Since ODR introduced this $\hat{\beta}\&\beta$ setup, a growing Behavioral Economics literature has attempted to elicit and estimate people's level of naivety in the time preference context.

Many behavioural economists followed these theoretical developments and started to empirically elicit and estimate these three parameters: present bias β, belief of present bias $\hat{\beta}$, and naivety. While the elicitation of the present bias has been harmonized through the introduction of the Convex Timing Budget (CTB) (Andreoni & Sprenger, 2012) technique, there is no unified method for the elicitation and estimation of the belief parameter. Going one step further, methods for measuring naivety i.e., a parameter measuring the relation between $\hat{\beta}$ and β, are still under exploration and development. While some studies explicitly estimated both $\hat{\beta}$ and β, some other works used various non-parametric ways to elicit the level of naivety. The literature on the topic is still at its infancy, with experimental designs and econometric techniques that vary substantially across papers.

Given the above, it seems interesting to try to understand both the syn-

thesis and the heterogeneity across these estimates. This thesis, similar to other meta-analytic projects, tries to summarize current empirical estimates of naivety. To be more specific, this thesis intends to answer the following two questions:

- First, what is the overall synthesis and variation in measured naivety across studies?
- Second, how do the different setup characteristics in each study affect the heterogeneity of estimates?

This thesis, together with the two research questions above, is largely inspired by Imai *et al.,* 2020, which gives a meta-analysis of the present bias estimates within the CTB context. To identify proper papers to analyze, a paper search following a six-step protocol was conducted. From 1,687 studies visited, ten studies, and forty-seven partial naivety estimates in total, are included in the dataset. I propose a consistent naivety measure across these ten papers and show how an approximate version of this measure could be constructed numerically. Due to the nested structure (some papers contain multiple dependent estimates), conventional meta-analytic models cannot be used, and alternatives must be found. As a result, random-effect-with-robust-variance-estimation (RE-RVE) model and random-effect-three-level (RE-3L) model were used to analyze the synthesis of the estimates. Five sets of explanatory variables, with its specification following Imai *et al.,* 2020 and Dellavigna & Pope, 2018, are included to study heterogeneity in these estimates.

This thesis finds the synthesis of present bias being 0.536 (0.558), the belief being 0.747 (0.793), and people understand 36.4% (32.1%) of their present bias (RE-RVE estimates shown the first and RE-3L estimates in the parenthe-

ses). Both methods yielded the present bias synthesis with enough statistical significance to reject the time consistency ($\beta = 1$) hypothesis. The naivety synthesis can also be shown to reject the hypothesis of sophistication. On the heterogeneity side, this thesis finds a significant difference between developed and developing countries, and components of study designs may also matter in the estimation.

Compared to Imai *et al.,* 2020 (28 papers with 220 estimates), this thesis had fewer estimates and, in turn, resulted in a small-sample issue. Reasonable efforts, especially twists in econometric methods, have been made to amend this issue, but the problem persists. Difficulties of this analysis are summarized in section 4.4, and methodological remedies are discussed in section 4.2.

The remainder of this thesis is structured as follows. Section 2 gives an overview of the relevant theories; Section 3 discusses the literature search methodology, naivety construction, and meta-analysis; Section 4 shows the data and results; at the end, section 5 concludes this thesis.

2 Review of The Theory

This section reviews theory relevant to partial naivety in present bias. For each agent, the partial naivety in present bias is a function of the parameters that measure the difference between her true underlying present bias parameter and her belief about it. Present bias is a time preference parameter measuring the difference between the agent's immediate and future valuations; this parameter is often associated with a quasi-hyperbolic discounting (QHD) inter-temporal model.

To begin with, present bias is deeply connected to time discounting issues, and probably one of the most famous anomalies of the conventional time discounting model. However, there are many more anomalies relevant to the discounting issue. Frederick *et al.,* 2002 provided a summary of these issues. The common anomalies of time discounting include: (1) the sign effect, in which gains are discounted more than losses; (2) the magnitude effect, in which small outcomes are discounted more than large ones; (3) the delay-speedup asymmetry, in which the discount rate can be drastically affected by whether the change of outcome time is framed as delayed or accelerated; (4) a preference for improving sequences, which discounts a factor's preference of improving a consumption sequence to diminishing counterparts, and many others. For these different anomalies, alternative models such as the self-awareness model (Pollak, 1968), the habit-formation model (Duesenberry *et al.,* 1949), and the reference-dependence model (Kahneman & Tversky, 1979) have also been proposed.

Out of all these alternative models, the quasi-hyperbolic discounting one that examines the present bias has the most widespread audience and has also become the "poster boy" of Rabin's PEEM criteria. As illustrated below, the hyperbolic discounting model only modifies one minor part of the conventional exponential discounting model yet fully captures a whole new source of variation and addresses one of the most frequently seen anomalies of the traditional model. The introduction of the quasi-hyperbolic model also popularized the discussion of the present bias parameter.

The motivation of the present bias study came from a widespread phenomenon

observed in real life: people often plan to do certain things but end up not doing them on time. In many practical cases, they systematically favour instant gratification or withhold immediate effort exertion at the expense of the future. These stories were also often interpreted as an issue of self-control or procrastination. For an agent who would contradict her own decision at a prior time, we say she is time-inconsistent. However, the standard exponential discounting model will always generate the opposite time-consistent behaviour, and thus, it will be impossible to address the self-control issue mentioned above.

2.1 Present Bias Parameter

Consider a sequence of time periods $\{0, ..., \tau, ..., T\}$. Firstly, we define that, for any τ, $U^\tau = \{u_\tau, ..., u_T\}$ is an inter-temporal sequence of agent's utility from period τ to T. Intertemporal choice theory also informed us that agents will have different valuations of U^τ in different periods, so I denote agent's view of U^τ as U_t^τ where $t \leq \tau$ is the period the agents stands at. In the exponential discounting model (EDM), the only time preference parameter is the exponential discounting parameter δ. That is, for an inter-temporal sequence of utilities, the EDM agent's utility function is:

$$U_t^\tau = \sum_{a=\tau}^{T} \delta^{a-t} u_a$$

with $\delta \in (0, 1]$ being the discount factor.

The quasi-hyperbolic discounting model, on the other hand, introduced one more time preference parameter β, which measures the immediate discounting between the current phase and the future. A QHD agent's inter-temporal

utility is

$$U_t^\tau = \begin{cases} \beta(\sum_{a=\tau}^{T} \delta^{a-t} u_a) & t < \tau \\ u_\tau + \beta(\sum_{a=\tau+1}^{T} \delta^{a-t} u_a) & t = \tau \end{cases}$$

with the new discount factor $\beta > 0$. In this case, $\beta \in (0, 1)$ shows the agent has a disproportionate preference for the immediate period against all future periods, whereas a $\beta > 1$ shows the opposite. Thus, β in this way measures an agent's bias for the present (or the future). As most of the people are believed to have a bias for the present rather than the future, β is frequently referred as the present bias parameter. As seen in equations above, a QHD agent coincides with a EDM one if $\beta = 1$.

This parameter, which introduces an imbalance between the immediate and the future into the inter-temporal utility function, was brought in to address the time inconsistency phenomenon observed in real life. We say that one agent's inter-temporal preference is *time-consistent* if for two inter-temporal sequences of utilities $U^t = \{u_t, ..., u_{t'}, ..., u_T\}$ and $\tilde{U}^t = \{\tilde{u}_t, ..., \tilde{u_{t'}}, ..., \tilde{u_T}\}$, and for any $t' \in \{t, ..., T\}$, we have

$$U_t^t > \tilde{U}_t^t \text{ if and only if } U_{t'}^t > \tilde{U}_{t'}^t$$

As explained in Halevy, 2015, time consistency requires agent's preference (represented by the utility function here) to be constant across periods. It implies that the agent will not have an incentive to deviate (or "regret") her prior choice in t. In other words, an agent's ex-ante preference (t) for the two given utility sequences will only depend on the utility sequences themselves but not on any ex-post t'. A deviation of time consistency can take the form of $U_t^t > \tilde{U}_t^t$ but $U_{t'}^t < \tilde{U}_{t'}^t$. This type of deviation is often denoted as the

"dynamic preference reversal" in literature. Observing the construction of EDM, it is easy to see that

Observation 1. *An EDM agent will always be time-consistent.*

In the mean time, a time-inconsistent agent's preference will depend on the time, and the introduction of β in the QHD model created a possibility to accommodate for this twist. One example is when $\beta \in (0, 1)$, people would systematically put a higher weight to t' when it is in the current period and would not do so when t' is in a future period. This type of time inconsistency is called *Presently Biased* preference.

Observation 2. *A QHD agent with $\beta \in (0, 1)$ will be time-inconsistent, and specifically, presently biased.*

Following O'Donoghue & Rabin, 1999, to illustrate this using a more compact method, I'll introduce a specific setup involving a task that has a deadline of T and requires a one-time effort provision to receive a payoff dependent on the effort provision time. That is, the inter-temporal utility sequence consists of two components: the payoff vector $\{v_1, v_2, ..., v_T\}$ and the effort cost vector $\{c_1, c_2, ..., c_T\}$, with both $v, c \geq 0$. For simplicity, we assume $\delta = 1$. And to show the potential problem of the presently biased agent, this setup will involve an immediate-effort cost exertion but a delayed reward payment. In this case, the inter-temporal utility an agent receives in period t if she decides to complete the task in period τ given $t \leq \tau$ is

$$U_t^\tau = \begin{cases} \beta v_\tau - \beta c_\tau & t < \tau \\ \beta v_\tau - c_\tau & t = \tau \end{cases}$$

with $\beta \in (0,1)$. And the agent coincides with an EDM agent if $\beta = 1$, that is

$$U_t^\tau = v_\tau - c_\tau \text{ if } t \leq \tau$$

As the EDM agent always faces the same optimization problem across all t, their decision will always be consistent. However, the higher weight for the current period, introduced by β, allows the agent to "change her mind" at different t. At time t, the agent will choose period τ to exert the effort if $\beta v_\tau - \beta c_\tau \geq \beta v_{\tau'} - \beta c_{\tau'}$ for any $\tau' \in \{t+1, ..., T\}$ and $\beta v_\tau - \beta c_\tau \geq \beta v_t - c_t$. However, when period τ arrives, this agent will change her idea if there exists $\tau' \in \{\tau+1, ..., T\}$ such that $\beta v_{\tau'} - \beta c_{\tau'} > \beta v_\tau - c_\tau$.

Intuitively, the present bias twists the optimization problem across different periods and causes people to contradict their prior decisions. Examples in real life include friends who say they will hit the gym in the next month but end up not doing it when the time arrives. In this case, we can easily predict that presently biased agents will likely delay the task. The scenario described above can be easily mirrored in the one-time immediate reward and delayed cost scenario, in which the presently biased agent will push the reward to earlier compared to the EDM ones. We can summarize these findings in the proposition below:

Observation 3. *A QHD agent with $\beta \in (0,1)$ will prefer a smaller immediate reward to a larger delayed award but will reverse their preference if both awards are shifted to the future by the same amount of time.*

2.2 Naivety, Sophistication and The Perceived Present Bias

After the framework of the present bias analysis was introduced, further research has shown that to understand agents' time-preference-related behaviour fully, we also need to know how much an agent understands her present bias; this motivates the study of the partial naivety in present bias. Partial naivety is a comparative measure between the present bias and the corresponding belief. Specifically, many model predictions crucially rely on the assumptions made regarding the agents' beliefs about their beta. The belief mentioned here is also often denoted as the *perceived present bias*.

A consensus has been reached to model time-inconsistent preferences as a game between "sub-agents" who are essentially the agent herself at different periods to understand their belief in her present bias. Each sub-agent will choose her strategy in her current period while her future selves would control the outcome.

The two key concepts, naivety and sophistication, were introduced here. When someone's belief about her present bias is denoted as $\hat{\beta}$, then she is sophisticated about her present bias if $\beta = \hat{\beta}$ and she is naive if $\beta \neq \hat{\beta}$. Furthermore, a person will be an absolute naïf if $\beta \neq 1$ but $\hat{\beta} = 1$.

Going back to the ODR setup:

$$U_t^\tau = \begin{cases} \beta v_\tau - \beta c_\tau & t < \tau \\ \beta v_\tau - c_\tau & t = \tau \end{cases}$$

again assuming $\delta = 1$ for simplicity. ODR studied behavioural differences between time-consistent, sophisticated, and fully naive agents and proposed a "perception perfect strategy" (PPS) to solve the optimum for these three types of agents. Their results are three-fold.

First, the naive and time-consistent agents use the same PPS solution concept, while the sophisticates have a different PPS. This informed us that a change in $\hat{\beta}$ might change the agent's behaviour. Second, full naifs procrastinate more than time-consistent people; this finding aligned with the introduction of the present bias parameter β. Third, sophisticates procrastinate weakly less than naifs; this finding captures a different force behind a present-biased agent's decision process. This shows that a sophisticated agent will have a better welfare outcome than the complete naif if earlier completion generates higher utility for the agent.

Latter empirical works have shown that behaviours of naifs are different to the sophisticates in contexts similar to above. Naive agents are more likely to be exploited in gym contracts and credit markets and less likely to optimize in contexts such as smoking and alcohol consumption fully. This set of studies raised a significant concern on how naivety in time preference can affect people's welfare in their life decision and led to a series of theoretical and empirical contributions about naivety.

One significant development of this topic, both theoretically and empirically, is to switch the focus from absolute sophisticates/naifs to partial naivety. This is in line with the publication of DellaVigna & Malmendier, 2004. Partial naivety assumes people's naivety level can be more various than the two most

polarized situations (completely sophisticated/naive) and opens a new door for empirical work to elicit people's specific naivety level.

The importance of partial naivety, in my opinion, is three-fold. First, it is reasonable to suspect that partial naifs will behave differently to sophisticates and complete naifs in usual time-preference-relevant contexts, as partial naifs understand parts of her present bias. Second, it is reasonable to believe that most people will be neither sophisticates nor complete naifs. Thus, a polarized assumption will offer us a biased estimate if the distribution of overall naivety level across the population is not symmetric to the halfway point. Last, the relationship between naivety and welfare leads economists to examine potential methods of intervention. One most frequently used—as Amador *et al.,* 2006 and Bryan *et al.,* 2010 studies that aim to reduce agents' welfare loss— is the commitment contract. In this case, a necessary condition for commitment demand is (at least partial) sophistication. Intuitively, a fully naive person will never realize that she has a self-control problem, so she will never seek the commitment to help herself as a consequence.

As this is a relevantly new field in Economics, recent theoretical and empirical contributions about partial naivety are made each year. One key issue of empirical works here lies in how to elicit agents' naivety parameters. As naivety is essentially a relative measure between the present bias and the perceived present bias, it seems natural to propose the following 3-step method to achieve this.

- Step 1: Elicit the agent's present bias parameter
- Step 2: Elicit the agent's perceived present bias parameter

- Step 3: Find a way to construct a naivety measure using the two parameters elicited above.

The 3-step method may sound a bit tautological, but the difficulty lies in working on these in the actual empirical setup. The following three subsections will discuss the three steps mentioned above and spoiler here: none of them is easy. Indeed, there are other ways to elicit and estimate the naivety parameter, including some non-parametric methods that do not explicitly elicit the two parameters. These methods will be further discussed in the methodology section, in which I propose a consistent measure across these naivety studies.

2.3 Elicitation of present bias parameter

Eliciting the present bias parameter is not an easy task. Due to the identification issue mentioned in Andersen *et al.,* 2008, eliciting it using experimental variations (Andreoni, Kuhn, *et al.,* 2013) seems to be the only feasible option. Moreover, Andersen *et al.,* 2008 showed that if the utility is assumed to be linear with the actual function being concave, the estimated coefficient will then be biased downwards. This finding provoked a new search of elicitation techniques, which resulted in the double multiple price list (DMPL) technique proposed by Andersen *et al.,* 2008 and the convex timing budget (CTB) technique proposed by Andreoni & Sprenger, 2012. Andreoni, Kuhn, *et al.,* 2013 made a detailed comparison between both methods.

The critical takeaway from Andersen *et al.,* 2008 is that they showed a joint elicitation of utility curvature and time preferences results in significantly lower discount rates than separate elicitation of discount rates. Thus, they argued that elicitation of risk preferences and time preferences should be conducted in two different stages.

The convex timing budget technique, proposed by Andreoni & Sprenger, 2012, is based on the DMPL method but simplifies this to a simple stage while guaranteeing successful identification of all three parameters concerned (utility curvature, exponential discounting, and hyperbolic discounting parameter). This technique is motivated by the observation that the agent would prefer interior solutions to corner solutions in a two-dimension allocation if being risk-averse. That is, any convex combination of allocations $(\$X, 0)$ and $(0, \$Y)$ should be preferred to them both. They used this argument and showed that parameters' identification could be achieved by convexifying the decision space. The CTB technique offered experiment subjects several allocation bundles, including corner solutions $(\$X, 0)$, $(0, \$Y)$, and convex combinations of them: $(1 + r)x_t + x_{t+k} = Y$ in which the interest rate $r = \frac{Y}{X} - 1$. CTB uses the slope variation to identify the utility curvature parameter and variation in k to identify both time preference parameters.

2.4 Elicitation of Perceived Present Bias

As described above, to understand time inconsistency, it is equally important to understand the true underlying present bias parameter β and level of naivety. With the knowledge of β, to understand naivety, the next step would be obtaining the corresponding perceived present bias $\hat{\beta}$ from the agent. For each agent, we need to elicit her first-order belief about her present bias β. This subsection will discuss how $\hat{\beta}$ is elicited in the literature.

It is hard to summarize how to elicit the perceived present bias. Unlike the present bias, the elicitation of which has been unified by the introduction of the CTB method, elicitation methods of $\hat{\beta}$ are still diverse. So, I will start with an

example of this elicitation, in which it has an experimental effort provision set up in line with one popular method in this field, which often involves designing an effort task that requires immediate effort exertion and gets delayed benefits.

In this method framework, people decide how to exert their effort in exchange for delayed rewards in a finite time frame at $t \in \{1, 2, ..., T\}$. All agents are assumed to have a QHD utility $U^t = u_t + \beta \sum_{\tau=t+1}^{T} \delta^\tau u_\tau$. In this case, the utility component consists of the vector (e, w), with e being the effort cost and w being a measure of the delayed reward. For example, in a given effort task setup, e could be how many tasks the agent chooses to do, and w could be the piece payment rate for each task completed.

A direct way to elicit belief is to ask participants to make statements about their belief. First, researchers are interested in $\hat{\beta}$, but apparently, they cannot ask participants to state the value of this parameter directly. There are at least two reasons behind it. Firstly, asking them directly may not guarantee the truthfulness of their reports.

Second, and more fundamentally, most participants will not have a well-structured belief about a quantitative parameter β. Thus, we need to derive and elicit their belief parameter from their behaviour. In this case, some researchers use participants' effort predictions for a future date as a surrogate for the perceived present bias. This is reasonable if (1) a truthful statement can be guaranteed and (2) it is indeed their perceived present bias that dictates people's prediction of their future exertion. The first point can be achieved by incentive mechanisms mentioned later in this section. For the second point, it is equivalent to state that we need a mapping from agent's belief $\hat{\beta}$, to the

agent's optimal e and then to her inter-temporal utility U^t. The usage of this method is seen in Augenblick & Rabin, 2019 and Fedyk, 2018.

In Augenblick & Rabin, 2019, participants were involved in an effort task setup, in which people choose e, that is, how many tasks to do within each period $t \in \{1, 2, ..., T\}$, and they would be paid ew at the end period T. The cost of e to each agent follows a cost function $C(e)$. For the belief elicitation purpose, subjects were asked to predict what they will do at time τ when it is $t < \tau$. The participant's utility takes a QHD form and thus the following function

$$\hat{\beta}\delta^{T-\tau}ew - C(e)$$

will dictate the agent's prediction about what she will do in the τth period in any $t < \tau$. Note that this is different from the agent's future decision at t about how much to work in τ. The utility of future decision is:

$$\beta\delta^{T-t}ew - \beta\delta^{\tau-t}C(e)$$

Intuitively, the prediction problem asks the participant to imagine she is already at period τ and decide in this imaginary scenario. Thus, it is $\hat{\beta}$ rather than β in the utility function; whereas the future decision asks the participant to decide what to do at τ at $t < \tau$, and thus both payoffs and effort exertion happen in the future. In this case, the e reported in the prediction problem is precisely the solution to the utility maximization problem based on e.

$$e^* = \text{argmax}_e \hat{\beta}\delta^{T-\tau}ew - C(e)$$

In this case, with sufficient exogenous variation in w, it is possible to recover $\hat{\beta}$ from observed prediction choices of e^*.

However, the maximization process above is only valid if the belief from participants is straightforward, that is, the belief would be a parameter not associated with agents' utilities. One common anomaly of this assumption, which also applies to Augenblick & Rabin, 2019, is that beliefs were distorted because participants were using it as a commitment device strategically. In this case, Augenblick & Rabin, 2019 showed that a "closed form" relationship between e^* and the utility can also be found.

The other problem with the belief statement, as mentioned earlier in the discussion, is how to ensure that people state their beliefs truthfully. This is often solved by incentivizing belief statements, for which the Becker-DeGroot-Marschak mechanism is frequently used to make sure the statement-revealing process is incentive-compatible.

The consensus among researchers is that incentive mechanisms can be used to ensure a truthful report of belief. However, specific incentive structures can also cause distortions in eliciting beliefs. Specifically, in Augenblick & Rabin, 2019, due to their extensive use of reminders, they need to imply the additional assumption of time-invariance, defined by Halevy, 2015 to sort the distortion caused by their accuracy bonus incentive structure for predictions. Time invariance means that for any pair of $(e, w; k_1)$ and $(e', w'; k_2)$ where k stands for the payment schedule, and any pair of period t, t', the following statement

must hold:

$$((e,w),t+k_1) \sim_t ((e',w'),t+k_2) \iff ((e,w),t'+k_1) \sim_{t'} ((e',w'),t'+k_2)$$

Intuitively, this means the preference between $(e,w;k_1)$ and $(e',w';k_2)$ of the agent will not depend on the period the agent makes the comparison; thus, her preference will be invariant across any t. Using words in Halevy, 2015, "If preferences are time invariant, only time relative to the evaluation period matters." This is certainly not a light assumption to make. It is also important to note that the assumption of time invariance should not be confused with time inconsistency. In this specific case, time consistency refers to $((e,w),t+k_1) \sim_t ((e',w'),t+k_2) \iff ((e,w),t+k_1) \sim_{t'} ((e',w'),t+k_2)$.

As seen above, Augenblick & Rabin, 2019 's elicitation also came at the expense of several strong assumptions. Generally, measuring beliefs will be a difficult task, and we need different assumptions to tackle them. On the other hand, different designs will lead to quite different estimates of beliefs. For example, the belief elicited can be about the agent herself or others (Fedyk, 2018). Different incentives should also have different effects on the beliefs elicited. One important goal of this thesis is to understand how these different assumptions, methodologies, and incentives affect the elicited value of the perceived present bias.

2.5 From the estimates to naivety

The final complexity is how to measure naivety once we have properly measured β and $\hat{\beta}$. There seem to be some straightforward and intuitive choices to propose a measure about how $\hat{\beta}$ is different/similar to β. For example,

$\hat{\beta} - \beta$ (DellaVigna & Malmendier, 2004, Bousquet, 2017) seems to be directly measuring the difference between two parameters. Besides, a ratio term $\frac{1-\hat{\beta}}{1-\beta}$ (Augenblick & Rabin, 2019) is an intuitive measure of how much the person actually understands her own bias. However, as discussed in Ahn *et al.,* 2019, both measures are problematic, and they showed that within the sets of naivety measured by the methods above, agents could make more sophisticated decisions despite being more naive in the parameter, and thus actions by the more naive individual do not reflect the naivety elicited by the constructed parameter mentioned above. In turn, they proposed two approaches as alternatives.

In the first approach, they measure the underdemand for commitment and determine the agent to be more naive if she rejects more beneficial commitments than others. In the second approach, agents are given a specific menu of choices. Agents are determined to be more naive if their excessive valuation of the menu (anticipated utility from the menu choice minus actual utility from the menu choice) is higher than others' excessive valuation.

Ahn *et al.,* 2019 's work offered another direct measure of eliciting naivety, but it seems to be difficult to align these estimates with the ones constructed more conventionally by methods mentioned earlier. The essence of a meta-nalysis is to include as many estimates while maintaining the possibility to compare them to each other consistently – a difficult balance to achieve, especially in the not-yet-unified context such as naivety. The following methodology section discusses my effort in making this happen.

3 Methodology

The methodology section builds upon the theory section and discusses how the paper selection (sample selection/creation), naivety variable construction (key variable), and meta-analytic methods were facilitated in this thesis.

3.1 Paper selection

The key for conducting a valid meta-analysis is to create a valid study sample following an objective and unambiguous inclusion protocol. The selection in this thesis aims to identify a broad and up-to-date sample of high-quality papers that examined the naivety level in the time preference context while also maintaining the feasibility of implementation. A six-step protocol is designed for the paper selection purpose.

3.1.1 Step 1

Similar to other meta-analysis work such as Imai *et al.,* 2020, the initial step is to identify an anchor paper as the essential reference. This reference will serve as a focal point of the selection process. The search will be based on its references and citations. After careful consideration, Handbook of Behavioral Economics (Bernheim, DellaVigna, *et al.,* 2019) was chosen for three reasons: First, it is recently published, so it will not omit too many relevant new studies in the context. Second, this project provides a comprehensive overview across sub-fields of Behavioral Economics, thus ensuring a good representation and breadth on the paper sample. Third, the Handbook focuses on important developments in Behavioral Economics, thus ensuring that the papers included in the dataset are of high quality.

3.1.2 Step 2

The second step of the paper search identifies relevant chapters from the handbook. This step aims to filter some irrelevant information from the handbook to make the search process more efficient. As stated before, it is important to make the search process as objective and unambiguous as possible. After a careful consideration, a word frequency check was conducted to identify each chapter's relevance. In this step, two word-related queries were conducted. The first query is based on time-preference-related words, which include "time-consistency", "present bias", "procrastination", and their variations. The second query uses naivety- and sophistication-related words and their variations. Then the chapters were ranked based on this word frequency check. Out of the eight chapters in the handbook (issue 1 plus the intertemporal choice chapter in issue 2), five chapters were chosen as the key reference for this selection process. The selected chapters are Inter-temporal Choices (Ericson & Laibson, 2019) Behavioral Household Finance (Beshears *et al.,* 2018), Behavioral Public Economics (Bernheim & Taubinsky, 2018), Behavioral Industrial Organization (Heidhues & Kőszegi, 2018), and Structural Behavioral Economics (DellaVigna, 2018). A summary of word frequency check in the selected chapters is shown in table 1[1].

[1]The check is conducted by summing word counts of all words listed in the word query (discussed in this part)

Table 1: Word Frequency Summary

Paper selected	Time preference	Sophistication and Naivety
Inter-temporal choice	117	76
Behavioral Finance	31	38
Behavioral Public Economics	90	83
Behavioral IO	29	420
Structural Behavioral	63	67

3.1.3 Step 3

Step 3 of the selection process searched the five selected chapters' references and citations [2]. I used Web of Science for the reference check and Google Scholar for the citation check. The search of references yields 1386 results, excluding all redundant results. And the citation search yields 295 results, with all redundancy excluded. Then I used Zotero to fetch all the 1681 items into a sheet containing both the publication records and the abstract.

3.1.4 Step 4

The fourth step of this process filters the 1681 papers based on their abstracts. This step aims to identify if this paper is relevant to this meta-analysis. Similar to step 2, an inclusion protocol based on different sets of words is used here for the filtering purpose. Five sets of keywords were proposed. The first set includes "time preference", "time inconsistency", "time-inconsistent", and "inter-temporal choice". The second set includes "procrastination", "procrastinate", "present bias", "quasi-hyperbolic", and "delay". The third set includes "sophistication", "sophisticated", "naive", "naïf", "naivete", "naivety", "perceived present bias", and "belief elicitation". The fourth set includes "commitment", "reminder", "self-control", "addiction", and "learning". The fifth set includes "experiment", "gym attendance", "smoking cessation", and

[2]Reference: papers that are included in the reference section of the five chapters mentioned above. Citation: Latter papers that cite one of these five chapters.

"tasks". Two inclusion protocols were set for this stage of selection:

First, a paper is included if at least one word in at least two sets appeared in either the title or the abstract (protocol 1). Second, a paper is included if at least one word in set 4 appeared in the title or the abstract (protocol 2).

Protocol 1 is set to rule out papers that only studied the present bias but not the naivety or vice versa. Protocol 2 is essentially an exemption from protocol 1 as many papers covering fieldwork of eliciting naivety lie in the context of commitments, self-control, addictions, and learning. By implementing this protocol and excluding redundant studies, 105 papers were selected from the 1,681 search results in the step before.

The five sets of keywords are not exhaustive lists of words. However, the key here is to do a feasible check on the vast amount of references, so it should be regarded as one feasible approach. Compared to the word frequency check in step 2, this is a more refined selection protocol because this step aims to filter out irrelevant papers caught in step 2. The first set of words contains keywords discussing generic time preference issues in the field. The second set puts focus on the exact present bias problem and its implications. The third set covers many variants of words relevant to sophisticates and naïfs. The fourth set covers important field works focusing on the implications of the naivety context. The last set contains frequent empirical setups examining both the present bias and the naivety context.

3.1.5 Step 5-6

The fifth step of selection is based on a quick examination of the 105 papers selected. The goal is to identify whether each paper provides at least one empirical section. This step rules out many theory-based papers from the final dataset. These theory papers are categorized into another collection as a valid theory reference. 49 papers were selected out of the 105 candidates in this process

The last selection step requires a detailed reading through all of the remaining papers and identifying if it is possible to build up an estimate of naivety from their reported results, and if so, what level of estimates can be obtained from the paper. This step mostly excluded empirical papers that assumed either complete sophistication or complete naivety. This step yields out ten papers left for our analysis.

The ten papers selected are Augenblick & Rabin, 2019, Fedyk, 2018, Carrera *et al.,* 2019, Bai *et al.,* 2017, Garon *et al.,* 2015, Wong, 2008, John, 2020, Yaouanq & Schwardmann, 2019, Acland & Levy, 2015, and Chaloupka IV *et al.,* 2019. A diagram of the full six-step paper selection process is shown in the figure below.

3.2 Naivety measure construction

The crucial measure in this meta-analysis is the one recording naivety reports in each paper. We need to construct a comparable measure to understand the synthesis and variation of estimates across studies, and it also serves as the dependent variable in the study characteristic heterogeneity analysis. I will first

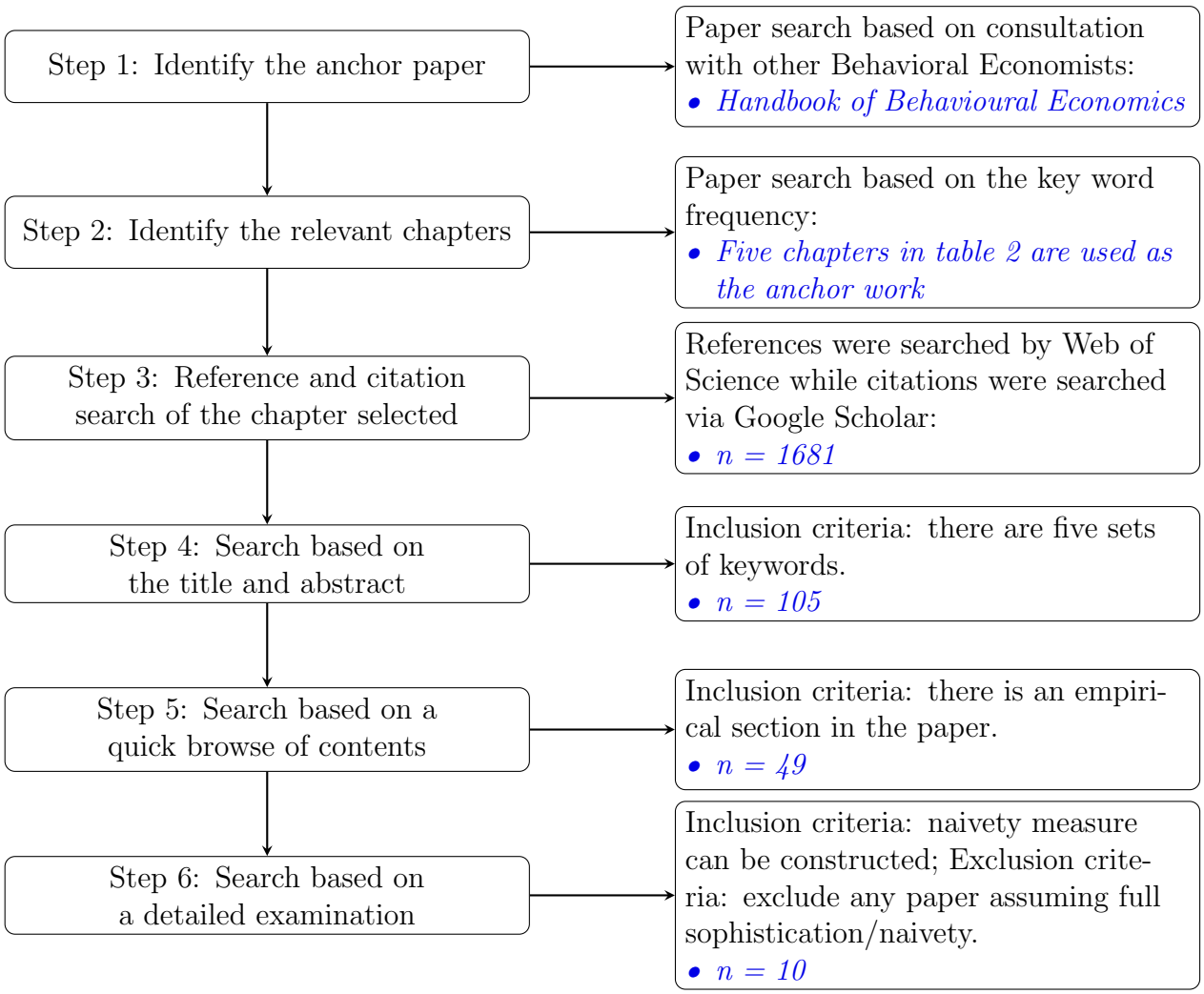


Figure 1: Paper Search Process

compare the estimates across papers included and then propose a consistent measure for naivety.

The first thing to understand about the naivety parameter is that it is a relative parameter measuring the distance between the time preference parameter β and the belief parameter $\hat{\beta}$. There are two ways to proceed from here. First, the naivety can be obtained by the 3-step method mentioned in Section 2.2. The first and second steps estimate both the time preference and the belief parameter using a structural model. I will refer to these estimates as structural ones due to this characteristic. The last step is to utilize specific post-estimation techniques, such as the delta method, to obtain the naivety parameter and its standard error. Acland & Levy, 2015, one study utilised this method, estimated parameters containing β and $\hat{\beta}$ structurally and then used delta methods to obtain an estimate of naivety.

On the other hand, in the ten papers selected, some naivety estimates are obtained without estimating the two parameters structurally. Consider the QHD model again, at period 0, the QHD agent's inter-temporal utility is

$$U(u_0, ..., u_T) = u_0 + \beta[\sum_{t=1}^{T} \delta^t u_t]$$

an agent will be a time-consistent person if $\beta = 1$. In this case, the distance between β and 1 measures the agent's time inconsistency. The $\hat{\beta}$ parameter, on the other hand, measures how much the agent understands his present bias. Specifically, the distance between $\hat{\beta}$ and 1 intuitively measures how much of the present bias the agent can anticipate, and the distance between $\hat{\beta}$ and β measures how much of the present bias the agent cannot predict. This

observation gives us a more "wholesome" measure of partial naivety. People's knowledge of their naivety level can be measured as:

$$\textbf{Naivety level} = \frac{\textbf{predicted inconsistency}}{\textbf{Total inconsistency}}$$

Intuitively, this measures how much the agent understands his naivety; Practically, since almost all the experimental partial-naivety papers involve both the actual outcome and people's statement of their future outcome, the ratio above can be easily calculated from observed data. This type of report is seen in several papers included in this meta-analysis, including Augenblick & Rabin, 2019, Fedyk, 2018, Wong, 2008, and Yaouanq & Schwardmann, 2019.

Going back to the Augenblick & Rabin, 2019 setup mentioned in Section 2.4, where the optimal effort chosen by an agent is

$$e^* = \text{argmax}_e \hat{\beta}\delta^{T-\tau} ew - C(e)$$

With $w, C(.), \tau$ fixed, e^*, the optimal effort chosen by the participant, is essentially a function of $\hat{\beta}$, the agent's belief of her present bias. Then the predicted inconsistency in this case is $e^*(1) - e^*(\hat{\beta})$, where $e^*(1)$ refers to the optimal effort of a time-consistent-version of the agent ($\beta = 1$). Similarly, total inconsistency here is $e^*(1) - e^*(\beta)$, where β is the agent's actual present bias. Utilising the naivety measure defined above, in Augenblick & Rabin, 2019 it is:

$$\textbf{Naivety level} = \frac{e^*(1) - e^*(\hat{\beta})}{e^*(1) - e^*(\beta)}$$

The numerator here measures the distance between the behaviour of a time-consistent version of agent and the behaviour of a "belief-version" of agent;

while the denominator measures the distance between time consistency and the behavior of the actual present bias level. From this interpretation, the analogue of this intuitive measure in the structural context can also be easily defined:

$$\textbf{Naivety level} = \frac{1-\hat{\beta}}{1-\beta}$$

where the numerator and the denominator have the exactly same interpretation. For those papers which reported the structural estimates of both the belief and the time preference parameters, this measure can be constructed even if they do not report the naivety estimate in this format. By this, we have a consistent proxy of the naivety estimate across different studies. This measure, however, is in no way near perfection, as critiqued by Ahn *et al.,* 2019, the shortcomings of this construction are denoted at the end of the prior section. The compromise made here is for the meta-analytic objective. Since most of the prior work was done in this manner, it does not make sense to use a different, new measure to synthesize them, albeit the new measure is a more comprehensive choice for measuring naivety itself.

For those studies which did not construct the naivety estimate using the method above but estimated $\hat{\beta}$ and β explicitly; the next step is to derive the naivety estimate from the current information. Notice that both the numerator and the denominator fall in the class of $Y = X + c$ where c is a constant. Thus, it does not change their respective variance, and their mean are changed to $1 - \mu_{\beta}$ and $1 - \mu_{\hat{\beta}}$ respectively. The next step is to deal with the ratio, as there are no closed-form solutions for the ratio change. The delta method, more specifically, a Taylor approximation, was used to find an approximate solution. The mean-variance vector $(\mu_{naivety}, \sigma^2_{naivety})$ can be approximated as

shown below.

$$\mu_{naivety} \sim \frac{1-\mu_{\hat{\beta}}}{1-\mu_{\beta}} - \frac{Cov(\beta,\hat{\beta})}{(1-\mu_{\beta})^2} + \frac{1-\mu_{\hat{\beta}}}{(1-\mu_{\beta})^3}\sigma_{\beta}^2$$

$$\sigma_{naivety}^2 \sim (\frac{1-\mu_{\hat{\beta}}}{1-\mu_{\beta}})^2[\frac{\sigma_{\hat{\beta}}^2}{(1-\mu_{\hat{\beta}})^2} + \frac{\sigma_{\beta}^2}{(1-\mu_{\beta})^2} - \frac{2Cov(\beta,\hat{\beta})}{(1-\mu_{\beta})(1-\mu_{\hat{\beta}})}]$$

The process of calculating the approximation can be found in the Appendix. The covariance item can be recovered by replicating regressions run in the original study. For this, the reports and the SEs of the time-preference parameter and the belief parameter are recorded, and the naivety proxy is constructed based on these observations. For those constructed from structural estimates, their approximate standard errors are also included. For the non-structural ones, the standard error (SE) obtained from observations will be zero. Luckily, for the few studies that need construction of the naivety estimate, the data needed was available.

3.3 Methodology for the naivety synthesis analysis

We can then step to the synthesis analysis once a consistent measure of naivety is constructed. As noted by Glass, 1976, meta-analysis is essentially "the statistical analysis of a large collection of analysis results from individual studies to integrate the findings." Suppose we have a set of m naivety estimates n_i, then the estimate of the overall effect is:

$$\bar{n} = \frac{\sum_1^m w_i n_i}{\sum_1^m w_i}$$

with w_i being the weight we assign to each study. The estimate $\bar{n}$ is referred as the *synthesis* in this thesis. There are two common models used in this type of analysis: the fixed-effect model and the random-effect model. A

fixed-effect model assumes that all observed studies are estimating the same treatment/specification effect, while a random effect model allows specifications of each estimate to be varied across the studies. The fixed-effect model, the starting point of meta-analytic literature, assumes that all studies share a fixed underlying effect size; in our case, say the fixed size is n_0. The observed effect under this assumption follows a normal distribution with a mean equal to n_0 and a variance equal to σ^2. For each n_i, that is:

$$n_i = n_0 + \epsilon_i$$

with $\epsilon_i \sim N(0, \sigma^2)$. The *overall effect* we intend to recover is n_0. In this case, it can be shown that the inverse of each n_i's variance v_i gives the most efficient weight to $\bar{n} = \frac{\sum_1^m w_i n_i}{\sum_1^m w_i}$ to recover n_0. That is, in a fixed-effect model, the weight given to each estimate is:

$$w_{i(FE)} = \frac{1}{v_i}$$

and the synthesis of naivety estimates takes the form shown below.

$$\bar{n} = \frac{\sum_1^m \frac{n_i}{v_i}}{\sum_1^m \frac{1}{v_i}}$$

The intuition behind this is that the only random error approaches 0 as sample size goes up, which suggests that the study with the larger sample will yield more precise estimates. In this case, it seems reasonable to give more weight to the more precise estimates. (Hedges & Vevea, 1997)

Due to the various participant pools and backgrounds in this analysis, a random-effect model seems to be a more appropriate option here. In a random-effect model, each study has an underlying true effect θ_i. Assume that θ_i is

normally distributed with mean equal to overall effect n_0 and variance equal to τ^2. Then we have

$$\theta_i = n_0 + \kappa_i$$

with $\kappa_i \sim N(0, \tau^2)$. Then each n_i is modelled as a realization from a normal distribution centered around θ_i.

$$n_i = \theta_i + \epsilon_i$$

with $\epsilon_i \sim N(0, \sigma^2)$. Combining both equations above, we have

$$n_i = n_0 + \kappa_i + \epsilon_i$$

That is, each n_i can be decomposed into the sum of the overall effect and two independently normally distributed error terms, with κ_i measuring the between-study error and ϵ_i measuring the within-study error.

The optimal weight for a random-effect model is also an inverse of variance.

$$w_{i(RE)} = \frac{1}{v_i + \tau^2}$$

where v_i is still the variance of estimate n_i and τ^2 is the between-study error variance. There exist several estimation methods of τ^2 but the most frequently used one here is

$$\hat{\tau}^2 = \frac{(m-1)\sum_1^m 1/v_i}{(\sum_1^m (1/v_i))^2 - \sum_1^m (1/v_i^2)}$$

by Higgins *et al.,* 2003. This representation of $\hat{\tau}^2$ is also referred to as the Q-variance due to its derivation started from the Q-statistic, a test of homogeneity

between estimates. Under the null hypothesis,

$$H_0 : n_1 = n_2 = ... = n_m$$

the test statistic

$$Q = \sum \frac{1}{v_i}(n_i - n_{\bar{FE}})^2$$

, where n_{FE} is the synthesis from the fixed-effect model, has an approximate chi-square distribution with (m-1) degrees of freedom. The Q stat here can be thought of as the total variance using the fixed-effect synthesis as the mean. Then the $\hat{\tau}^2$ is calculated as an excess variance by:

$$\hat{\tau}^2 = \frac{Q - df}{C}$$

with df being degrees of freedom equal to $m - 1$ and C being a scaling factor that ensures v_i and $\hat{\tau}^2$ are on the same scale.

$$C = \sum(1/v_i) - (\sum(1/v_i)^2)/\sum(1/v_i)$$

Putting all the components together, the synthesis of naivety estimate within a random-effect model is:

$$\bar{n} = \frac{\sum_1^m w_{i(RE)} n_i}{\sum_1^m w_{i(RE)}}$$

with

$$w_{i(RE)} = \frac{1}{v_i + \frac{(m-1)\sum_1^m 1/v_i}{(\sum_1^m (1/v_i))^2 - \sum_1^m (1/v_i^2)}}$$

As discussed, Q-stat gives us a quick judgment of whether the estimates are homogeneous. However, it does not give us a quantitative measure of how heterogeneous these estimates are. For this, Higgins *et al.,* 2003 proposed the

I^2 statistic:

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \tilde{v}^2}$$

with $\tilde{v}^2$ being a chosen overall measure of v_i across studies, one example being $\bar{v}^2$ the average variance. I^2 here measures a relative relation between the between-study variance and the within-study variance. In the random effect model, this test statistic can be shown as

$$I^2 = \frac{Q - df}{Q}$$

with all terms defined as above. Intuitively, the *df* term measures variance from sampling error, and this given ratio measures how much of the total variance comes the from heterogeneity between studies.

The idea of random effect models relaxed the strict assumption of one single fixed effect across all studies, but it is still not applicable to studies with so-called "*nested estimates*." Nested estimates mean that we have more than one estimate reported in some of the studies. Those estimates in the same study are likely to be dependent, but both FE and RE models assumed estimates to be independent of each other. If we mistakenly treated dependent estimates as independent, studies that reported multiple effect sizes would automatically get higher weights. Two methods have been proposed to remedy this.

Hedges, Tipton, *et al.,* 2010b proposed the robust variance estimation (RVE) method to deal with dependent estimates nested within each study. This model has two specifications, the first being an analogue of GLS estimated by finding a feasibly optimal estimator, which is well-documented in their paper. The

second case, a hierarchical effects model, showed that the dependence problem could be bypassed by using the following weight for each estimate in a given study j.

$$w_{ij} = \frac{1}{v_{ij} + \omega^2 + \tau^2}$$

where v_{ij} is the variance associated with each estimate, ω^2 is the variance captured within a study cluster j, and τ^2 is the between-study variance. Then they used the method of moments to jointly estimate ω^2 and τ^2.

The estimate of τ^2 is quite similar to τ^2 in $w_{i(RE)}$. First, we calculate the Q-statistic in this class of model

$$Q = \sum_i \sum_j w_i n_{ij}^2 - \frac{(\sum_i \sum_j w_i n_{ij})^2}{\sum_i \sum_j w_{ij}}$$

Then τ^2 is estimated by

$$\hat{\tau}^2 = \frac{Q - (m-1)}{\sum\sum w_{ij} - (\sum\sum w_{ij}^2)/(\sum\sum w_{ij})} - \hat{\omega}^2$$

where $\hat{\omega}^2$ is the estimate of ω^2. The $\hat{\omega}^2$ here is estimated using a more complex construction, which is shown in the appendix.

The RVE model was shown by Hedges, Tipton, *et al.,* 2010a to have desirable asymptotic properties. For small-sample studies, Tipton, 2015 extended this model, and a small-sample adjustment method was proposed to ensure the RVE estimator is equipped with better small sample properties.

The second model proposed for nested estimates is a random-effect-3-level model (RE-3L). First, notice that the usual random-effect model can be con-

sidered a 2-level model.

$$\text{Level 1: } n_i = \theta_i + \epsilon_i$$

$$\text{Level 2: } \theta_i = n_0 + \kappa_i$$

To control the dependence between estimates within a study, we can add one more level controlling the clustering effect.

$$\text{Level 1: } n_{ij} = \lambda_i + \epsilon_{ij} \textit{ Participant level}$$

$$\text{Level 2: } \lambda_i = \theta_i + \kappa_{i2} \textit{ Cluster level}$$

$$\text{Level 3: } \theta_i = n_0 + \kappa_{i3} \textit{ Study level}$$

An example of the 3-level model is shown in figure 2. In this example, we have three studies (the second row) and five estimates (two from study 1, one from study 2, and two from study 3). We are interested in estimating n_0. Assume all the error terms associated with the 3-level model are normal, and then the distribution hierarchy takes the following form:

- n_0, mean of the level-3 distribution is the overall effect we want to estimate.
- In level-2 and for each study, the true effect size is a realization of the level 3 distribution by the assumption of the random-effect model. Then $\theta_1, \theta_2, \theta_3$ each is a realization of the level-3 distribution and the mean of the level-2 distribution, $\theta_i = n_0 + \kappa_{i2}$
- For each estimate in each study, the 3L model assumes it has its own true effect size as a realization of the study effect size. We denote these effect sizes as $\lambda_1, \lambda_2, \lambda_3$. Similarly, each λ is a realization of the distribution in level 2 and $\lambda_i = \theta_i + \kappa_{i3}$. For study 2, $\lambda_2 = \theta_2$ as there's only one

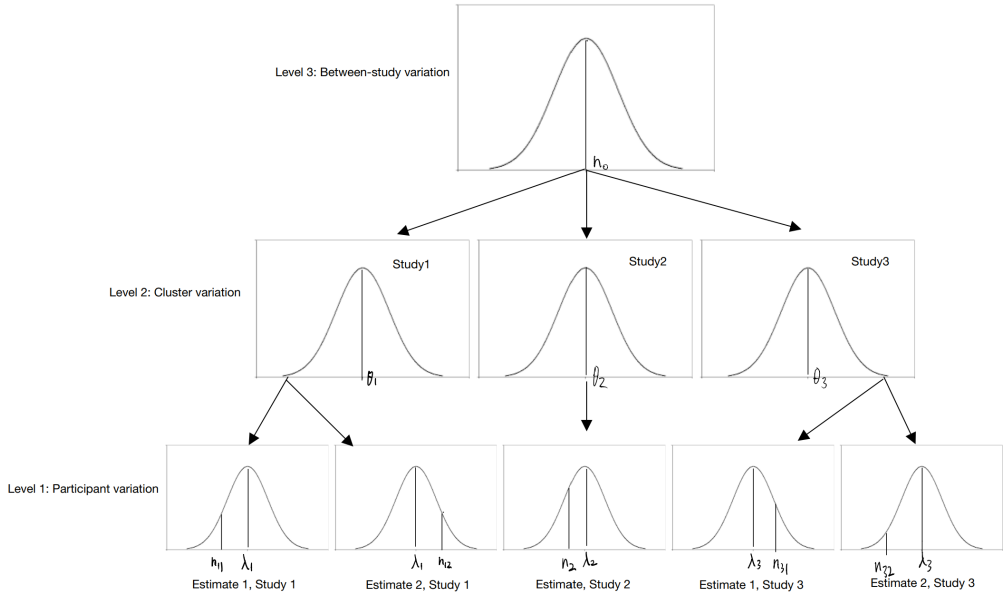


Figure 2: RE-3L: example

estimate from the study.

- Last, For the five estimates $n_{11}, n_{12}, n_2, n_{31}, n_{32}$, each is a realization from the level-1 distribution, with $n_{ij} = \lambda_i + \epsilon_{ij}$

Same to the two-level model, the three-level model can also be collapsed as one single equation:

$$n_{ij} = n_0 + \kappa_{i2} + \kappa_{i3} + \epsilon_{ij}$$

n_{ij} stands for the estimate reported from the jth estimate from the ith study in the dataset. The three error terms, κ_{i2}, κ_{i3}, and ϵ_{ij}, measure the between-study, within-cluster, and participant-level variation, respectively. Similar to the two-level model, we assume that three errors are at least mean-independent

to each other. This then implies:

$$Cov(\kappa_{i2}, \kappa_{i3}) = 0$$
$$Cov(\kappa_{i2}, \epsilon_{ij}) = 0$$
$$Cov(\kappa_{i3}, \epsilon_{ij}) = 0$$

Konstantopoulos, 2011 then showed that the conditions above implied that:

$$E[n_{ij}] = n_0$$

which indicates the mean of estimate effect sizes is the overall effect we want to estimate. Van den Noortgate *et al.,* 2013 showed that the class of RE-3L models can be estimated by the maximum likelihood method. Cheung, 2013 further suggested that the restricted maximum likelihood estimator (REML) is the default estimator for this class of analysis.

Test-statistic-wise, the RE-3L model shares many statistics with its two-level sibling. For example, the Q-stat is still valid for testing the homogeneity of estimates in a three-level model. However, the equivalence $I^2 = \frac{Q-df}{Q}$ does not hold for this class of model anymore. In order to account for the newly introduced level, the I^2 statistic in a three-level model is revised as:

$$I^2_{(2)} = \frac{v(\kappa_2)}{v(\kappa_2) + v(\kappa_3) + \tilde{v}}$$
$$I^2_{(3)} = \frac{v(\kappa_3)}{v(\kappa_2) + v(\kappa_3) + \tilde{v}}$$

with $v(\kappa_2), v(\kappa_3)$ standing for the variance of two error terms, and $\tilde{v}$ still being the chosen overall measure of v_i. The interpretation of the I^2 statistic here remains the same.

3.4 Heterogeneity analysis methodology

3.4.1 Econometric model

As discussed in the introduction, the second goal of this analysis is to see if different setups and specifications in each project vary the naivety estimate systematically. To study this, consider the following simple econometric model:

$$n_{ij} = \gamma_0 + \gamma_1 \boldsymbol{x_i} + \gamma_2 \boldsymbol{y_{ij}} + \psi_{ij}$$

in which $\boldsymbol{x_i}$ stands for different specifications across the papers and $\boldsymbol{y_{ij}}$ stands for different specifications across each estimate of each paper.

3.4.2 Explanatory variables in the heterogeneity analysis

The sets of explanatory variables are constructed to analyze how each paper's characteristics and treatments affect the reported estimates. The variables selected can be categorized into several groups.

The first group of variables contains basic information about each paper. It records the paper's name, publication status, and publication journal if the paper is published. The group name variable serves as the group ID for creating the summary statistics, and the status of publication is used as a control variable in the heterogeneity analysis.

The second group of variables records the characteristics of the sample. As demonstrated in Dellavigna & Pope, 2018, samples with different demographic characteristics may result in different prediction outcomes. These group variables record the sample size, the recruitment method, country demographics, and whether the sample comprises students or a mixed population.

The third group of variables records the instruments used in the study. Different elicitation forms may result in unanticipated outcomes, and this group variable records if a lab component, a survey, an online component, or a field task is involved in the study. The same task involved in each study is also recorded in this variable set.

The fourth group of variables recorded information about the naivety estimates. It records if the reported estimate is a structural one or not and whether it is an individual-level estimate or an aggregate-level estimate. It also includes the number of observations in the estimation; note that this is different from the sample size recorded in the second group as participants usually made more than one decision in the process.

The last set of variables denotes how the naivety is elicited. Several specific elicitation techniques are noted in this set: usage of the calendar when eliciting prediction, use of reminders (about agents' predictions), and the wording of the elicitation questions are included in this set. The wording of elicitation here refers to the question eliciting beliefs about the future, which verb the researcher used. For example, in an effort task prediction, a researcher may naturally ask: Can you "predict" how many tasks you will do in the next week? However, the word "predict" has several close substitutes such as "guess", "expect", and "think". The inclusion of this set of variables aims to understand if different elicitation techniques would affect the elicitation outcomes.

Altogether, the dataset includes five sets of variables plus information about the present bias, belief, and naivety estimates. A handbook of data coding is

included in the Appendix.

4 Data and Results

4.1 Overview of dataset

The dataset includes ten papers selected from the search process explained in the section above. All the reported estimates in these papers are included in this dataset, giving us forty-seven estimates.

Out of the ten studies included, six explicitly elicited and reported estimates of both the present bias parameter and the belief parameter, while the other four constructed the naivety measure in various ways. There exist several different regression types across studies. Most importantly, there exist both aggregate and individual estimates. Aggregate estimates here refer to pooling all data together and finding the aggregate estimate, while individual estimates stand for regressing on each individual's data and then using a measure to harmonize the finding across individuals, such as mean or median. The amount of individual estimates is about the same as the aggregate estimates. Out of the 22 individual estimates, 17 report the mean, while the other five report the median. Most of the estimates focused on eliciting beliefs about themselves. Fedyk (2019) is the only work that also elicited participants' beliefs about others.

Only 4 out of 10 studies included in this survey recruited participants from a university-based lab, and only 3 out of 10 studies have a university-only sample. This means the 10 studies selected here are more field-oriented compared to the similar former analysis by Imai *et al.,* 2020 (with an approximately 1:1

lab-to-field ratio). Note that mixed populations and university populations add up to 11. This is because one study Fedyk, 2018 included two types of experiments, one with students and one with a mixed population. Many studies included here used multiple components in their experiments, such as a mix of lab components and online tasks. The most frequent belief-eliciting wordings are "predict" and "think," and each stands for around 40% of the estimates.

Table 2: Summary Statistics

Variable	Frequency (Study)	Proportion (Study)	Frequency (Estimates)	Proportion (Estimates)
Estimates reported:				
Naivety estimates reported	10	100	47	100
PB estimates reported	6	60	38	80.9
Perceived PB estimates reported	6	60	38	80.9
Aggregate level	4	40	21	44.7
Individual level	7	70	22	46.8
Experiment types:				
Lab experiment	4	40	26	55
Field experiment	6	60	20	42.6
Survey	8	80	29	61.7
Online experiment	4	40	26	55.3
Elicitation:				
Elicitation: Self	9	90	42	89.3
Elicitation: Other	1	10	5	10.6
Belief wording: Predict	3	30	18	38.3
Belief wording: Think	2	20	19	40.4
Belief wording: Others	5	50	10	21.3
Recruitment:				
Recruitment: mixed-population	8	80	42	89.3
Recruitment: university	3	30	5	10.6
Study:				
Total number of studies/estimates	10	100.0	47	100.0

There are several ways for this dataset to be expanded. Observation-wise, the dataset can be expanded in two ways. First, different combinations of keyword searches can be conducted on Google Scholar. Second, even though the estimates reported by each paper have all been included in this work, it is still possible to decompose the original working data into specifications that researchers have not reported in their work and generate new estimates. For example, in Augenblick & Rabin, 2019, a bonus system was introduced to incentivize participants' truthful predictions. Then their estimates can also be structured in a bonus variation category, which they did not explicitly report in their work.

Variable-wise, the current dataset lacks a systematic measure of the incentivization method utilized in each study. Elicitation of naivety is more complex than elicitation of the time preference parameter because it involves another component of the belief elicitation, and the incentive methodology is crucial in eliciting the belief components. Overall, the target of the variable addition is to identify more sets of variables that directly affect both the estimates of the time preference parameter and the belief parameter.

4.2 Synthesis of the estimates

In this section, we will discuss the results of the synthesis analysis. As noted in the methodology section above, two methods (RE-RVE and RE-3L) were used to estimate the overall level of naivety in time preference. As mentioned in the methodology section, two characteristics of this data set motivated me to use these two specific methods. First, the true parameter value may not be consistent across all studies included due to the contrasting backgrounds of each study. For this, the random effect (RE) model is preferred to the fixed effect

(FE) model, as the latter assumes fixed true-parameter values across studies. Second, since multiple estimates are included in one study, these within-study estimates are not independent of each other, and this nested data structure violates the assumption of independence for the random effect model. For this, the two most conventionally used methods are indeed RE-RVE and RE-3L, as described in the methodology section above. The results from both models are shown below.

In both panels, the first two rows report the average estimates and the

Table 3: Synthesis Results

Panel 1: Synthesis from RE-RVE			
	Time preference	Belief	Naivety
Average	0.536	0.747	0.364
Confidence Interval	$[0.161, 0.911]$	$[0.378, 1.110]$	$[0.079, 0.648]$
τ^2	0.048	0.038	0.102
I^2	50.993	44.980	76.258
Observations	37	37	37
Studies	6	6	6
Panel 2: Synthesis from RE-3L			
	Time preference	Belief	Naivety
Average	0.558	0.793	0.321
Confidence Interval	$[0.305, 0.811]$	$[0.572, 1.010]$	$[0.075, 0.566]$
$v(\kappa_2)$	0.000	0.000	0.007
$v(\kappa_3)$	0.076	0.234	0.073
Observations	37	37	47
Studies	6	6	10

associated confidence interval. The test statistic on the following row (rows) is calculated from the model, and it gives the amount of heterogeneity given by the error specified in the model. Specifically, τ^2 represents between-study variation. The RE-RVE model was run in R by the *Robumeta* package by Fisher & Tipton, n.d., while the RE-3L was run in R by the *metafor* package (Viechtbauer, 2010). Due to the small sample size of this study, the estimates

given by the RE-RVE model were also adjusted according to Hedges, Tipton, *et al.,* 2010b.

Firstly, I will discuss results from the RE-RVE model. The first column, Time preference, reports the meta-analytic average of the present bias estimate. The meta-analytic mean here is 0.536. Comparing this with results in Imai *et al.,* 2020, the mean of present bias in this study sample is more severe than the mean in Imai *et al.,* 2020 's sample (between 0.88 and 0.96). However, the results here aligned with Imai *et al.,* 2020 's, as both studies reject the null of time consistency ($\beta = 1$) at conventional levels.

The τ^2 statistic here shows the variance component obtained from the between-study error. On the other hand, participant belief was estimated to be 0.747, which means the average participant belief across all studies included is around 3/4, suggesting most participants in studies do have at least a partial understanding of their bias. The average naivety is estimated to be 0.364, suggesting an average 36% understanding of their own bias. On a 5% significance level, the lower bound of the confidence interval is quite close to 0, which shows a higher severity of their misunderstanding.

The τ^2 statistic here shows the variance component obtained from the between-study error. The I^2 as mentioned in the methodology section is a percentage measure.

Panel 2 reports result from the RE-3L model. Two things are worth noting here. First, the results here are estimated using the restricted maximum likelihood method (REML). This is one of the most typical estimation meth-

ods used in meta-analysis for being easy to conduct. There may exist other feasible estimation methods here. Second, the results here are corrected by the Knapp-Hartung adjustment, which is shown by IntHout *et al.,* 2014 to provide more precise estimates. The two middle rows of the table report the variance contributions from the two errors. Here, $v(\kappa_2)$ stands for the cluster-level variation and $v(\kappa_3)$ stands for the between-study variation.

The syntheses of both β and $\hat{\beta}$ are about the same in the RE-3L model. In this small sample, $v(\kappa_2)$ is very close to zero, showing that the estimates within each study are highly similar, most likely due to the small number of studies included and the fact that estimates are clustered in the two studies included. The naivety synthesis, however, shows to be slightly less severe compared to the RE-RVE synthesis. The RE-3L model estimates that the average understanding of the present bias is about 32%.

4.3 Heterogeneity of estimates

The second research question focuses on how studies' experiment, treatment, and specification variations affect the naivety estimate. Here, I report the estimates by adding different explanatory variables group by group. The regression result is shown in table 4. The results reported are estimated by a linear regression allowing random effects across the studies.

The first column reports regression results from the sample characteristics. The three variables are all coded as binary ones. The general sample variable is equal to one if not all participants are university students. The sample country variable is equal to 1 if the sample is taken in a developed country. The lab recruitment variable equals 1 if the sample is recruited from a university-based

lab participant pool. As seen in the reported estimates, only the recruitment variable is weakly significant in this specification.

The second specification introduces a set of variables about components of the study. The lab, online, and survey components are binary variables equal to 1 if the study involves this component. This specification shows that both the sample country and the survey component are at least significant on a 5% level. The negative value of the sample country estimate suggests that naivety elicited in developed countries is significantly lower than naivety elicited in developing countries. However, the significance disappears after introducing other variables.

The third group of variables being added to the analysis is information on the estimated level. The two binary variables specify whether the estimate is an individual one or a structural one. Both of these two estimates are not statistically significant and do not change the prior estimates much. The sample country estimate was no longer significant after the introduction of these two variables.

There is one more set of variables about the elicitation information that has not been included. The binary variable remainder equals 1 if researchers explicitly remind participants of their prior predictions about this period when they are making a current decision. The "think and predict" variable denotes the wording used in the elicitation as a three-fold categorical variable. Due to the lack of estimates, the specification including this set of variables reported non-interpretable large coefficients and thus is omitted from the report here.

Table 4: Heterogeneity results

	Dependent variable:		
	Naivety		
	(1)	(2)	(3)
Base category on the RHS			
Sample: Developed vs developing countries	−0.112	−0.345***	−0.172
	(0.121)	(0.115)	(0.135)
Sample: General vs student	−0.201	0.210	0.274
	(0.123)	(0.175)	(0.171)
Recruitment: Lab vs other	−0.235**	0.304	0.232
	(0.090)	(0.223)	(0.272)
Experiment component: Lab vs no lab		−0.416	−0.278
		(0.306)	(0.352)
Experiment component: Online vs no online		0.066	−0.069
		(0.220)	(0.232)
Experiment component: Survey vs no survey		0.331***	0.356***
		(0.076)	(0.089)
Estimates: Structural vs non-structural			−0.102
			(0.089)
Estimates: Individual vs aggregate			0.145
			(0.096)
Constant	0.705***	0.190	0.020
	(0.128)	(0.167)	(0.212)
Observations	47	47	43
R^2	0.305	0.552	0.625
Adjusted R^2	0.257	0.485	0.537
Note:			*p<0.1; **p<0.05; ***p<0.01

4.4 Difficulty in comparing estimates

This meta-analysis project conducted a literature search over one-thousand six-hundred research research papers but identified only 10 papers suitable for analysis. This process alone shows the difficulty in making a consistent comparison across partial naivety estimates. The difficulty of this lies in the following issues:

First, there exist very few pieces of work associated with this topic. Compared with present bias studies, there were fewer works that empirically examined the partial naivety issue. Imai *et al.,* 2020 managed to find twenty-eight articles with which to work, with two hundred twenty estimates reported in their work. It is natural to expect that fewer pieces of work can be found for the partial naivety issues in the same context.

Second, within this limited body of work, naivety is measured in very different ways, making comparisons between studies difficult. Early researches in this field either focused on the theory side or did not measure $\hat{\beta}$ empirically.

Last, even within the set of estimates that had been empirically examined, further data is required to offer a valid comparison between them. For those estimates that were not directly transformable to the consistent measure proposed in the last section, no analysis can be made without the acquisition of further data.

5 Conclusion

Motivated by Imai *et al.,* 2020, this thesis provides a quantitative meta-analysis on partial naivety estimates in the time preference context. Ten studies and forty-seven estimates in total are included in this analysis. Both the synthesis and the heterogeneity issues have been studied in this thesis.

This thesis firstly re-visited theories relevant to both the present bias and its corresponding belief parameter. Both the theoretical developments and the elicitation techniques in the past twenty years have been discussed. Then, a six-step literature search method was proposed to identify relevant papers.

This thesis also provided a quick overview of meta-analytic methods and explained why the RE-RVE and RE-3L models were used to study the synthesis. The heterogeneity issue in this thesis, following Imai *et al.,* 2020, used a conventional linear regression model with observation and group-level variables.

Given the small data set, this thesis offers a glimpse of the naivety analysis synthesis in the literature. The synthesis here can be expanded and will be made more accurate by adding more available studies. This thesis completes its task by offering a valid estimate of the synthesis and also studying the heterogeneity issue.

However, the greater objective, similar to other meta-analytic works, is to try offering some insight about which is the best method of eliciting and estimating the naivety parameter. That challenge is indeed a fascinating albeit demanding one, but this thesis cannot address that challenge.

References

1. Acland, D. & Levy, M. R. Naiveté, projection bias, and habit formation in gym attendance. *Management Science* **61.** Publisher: INFORMS, 146–160 (2015).

2. Ahn, D. S., Iijima, R., Le Yaouanq, Y. & Sarver, T. Behavioural characterizations of naivete for time-inconsistent preferences. *The Review of Economic Studies* **86.** Publisher: Oxford University Press, 2319–2355 (2019).

3. Amador, M., Werning, I. & Angeletos, G.-M. Commitment vs. flexibility. *Econometrica* **74,** 365–396 (2006).

4. Andersen, S., Harrison, G. W., Lau, M. I. & Rutstrom, E. E. Eliciting risk and time preferences. *Econometrica* **76.** WOS:000255137300005, 583–618. ISSN: 0012-9682 (May 2008).

5. Andreoni, J., Kuhn, M. A. & Sprenger, C. *On measuring time preferences* tech. rep. (National Bureau of Economic Research, 2013).

6. Andreoni, J. & Sprenger, C. Estimating Time Preferences from Convex Budgets. *American Economic Review* **102.** WOS:000312093000008, 3333–3356. ISSN: 0002-8282 (Dec. 2012).

7. Augenblick, N. & Rabin, M. An Experiment on Time Preference and Misprediction in Unpleasant Tasks. *Review of Economic Studies* **86.** WOS:000493306500002, 941–975. ISSN: 0034-6527 (May 2019).

8. Bai, L., Handel, B., Miguel, E. & Rao, G. Self-control and demand for preventive health: Evidence from hypertension in India. *Review of Economics and Statistics.* Publisher: MIT Press, 1–55 (2017).

9. Bernheim, B. D., DellaVigna, S. & Laibson, D. *Handbook of Behavioral Economics-Foundations and Applications 2* (Elsevier, 2019).

10. Bernheim, B. D. & Taubinsky, D. Behavioral public economics. *Handbook of Behavioral Economics: Applications and Foundations 1* **1.** Publisher: Elsevier, 381–516 (2018).

11. Beshears, J., Choi, J. J., Laibson, D. & Madrian, B. C. in *Handbook of Behavioral Economics: Applications and Foundations 1* 177–276 (Elsevier, 2018).

12. Bousquet, L. *Measuring time preferences and anticipation: a lab experiment* tech. rep. (Citeseer, 2017).

13. Bryan, G., Karlan, D. & Nelson, S. Commitment devices. *Annu. Rev. Econ.* **2,** 671–698 (2010).

14. Carrera, M., Royer, H., Stehr, M., Sydnor, J. & Taubinsky, D. *How are preferences for commitment revealed?* tech. rep. (National Bureau of Economic Research, 2019).

15. Chaloupka IV, F. J., Levy, M. R. & White, J. S. *Estimating biases in smoking cessation: Evidence from a field experiment* tech. rep. (National Bureau of Economic Research, 2019).

16. Cheung, M. Modeling Dependent Effect Sizes With Three-Level Meta-Analyses: A Structural Equation Modeling Approach. *Psychological methods* **19** (July 2013).

17. DellaVigna, S. in *Handbook of Behavioral Economics: Applications and Foundations 1* 613–723 (Elsevier, 2018).

18. DellaVigna, S. & Malmendier, U. Contract design and self-control: Theory and evidence. *The Quarterly Journal of Economics* **119,** 353–402 (2004).

19. Dellavigna, S. & Pope, D. What Motivates Effort? Evidence and Expert Forecasts. *Review of Economic Studies* **85.** WOS:000428940700010, 1029–1069. ISSN: 0034-6527 (Apr. 2018).

20. Duesenberry, J. S. *et al.* Income, saving, and the theory of consumer behavior (1949).

21. Ericson, K. M. & Laibson, D. in *Handbook of Behavioral Economics: Applications and Foundations 1* 1–67 (Elsevier, 2019).

22. Fedyk, A. Asymmetric naivete: Beliefs about self-control. *Available at SSRN 2727499* (2018).

23. Fisher, Z. & Tipton, E. robumeta: an R package for robust variance estimation in meta-analysis. en, 16.

24. Frederick, S., Loewenstein, G. & O'donoghue, T. Time discounting and time preference: A critical review. *Journal of economic literature* **40,** 351–401 (2002).

25. Garon, J.-D., Masse, A. & Michaud, P.-C. Health club attendance, expectations and self-control. *Journal of Economic Behavior & Organization* **119,** 364–374 (2015).

26. Glass, G. V. Primary, secondary, and meta-analysis of research. *Educational researcher* **5,** 3–8 (1976).

27. Halevy, Y. TIME CONSISTENCY: STATIONARITY AND TIME INVARIANCE. *Econometrica* **83.** WOS:000349962900012, 335–352. ISSN: 0012-9682 (Jan. 2015).

28. Hedges, L. V. & Vevea, J. L. A Study of Equating in NAEP. NAEP Validity Studies. (1997).

29. Hedges, L. V., Tipton, E. & Johnson, M. C. Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. en. *Research Synthesis Methods* **1.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.17 164–165. ISSN: 1759-2887. `https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.17` (2021) (2010).

30. Hedges, L. V., Tipton, E. & Johnson, M. C. Robust variance estimation in meta-regression with dependent effect size estimates. en. *Research Synthesis Methods* **1.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.5, 39–65. ISSN: 1759-2887. `https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.5` (2021) (2010).

31. Heidhues, P. & Kőszegi, B. Behavioral industrial organization. *Handbook of Behavioral Economics: Applications and Foundations 1* **1,** 517–612 (2018).

32. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *Bmj* **327.** Publisher: British Medical Journal Publishing Group, 557–560 (2003).

33. Imai, T., Rutter, T. A. & Camerer, C. F. Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets. *Economic Journal.* Publisher: Royal Economic Society, Art–No (2020).

34. IntHout, J., Ioannidis, J. P. & Borm, G. F. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC medical research methodology* **14,** 1–12 (2014).

35. John, A. When commitment fails: evidence from a field experiment. *Management Science* **66.** Publisher: INFORMS, 503–529 (2020).

36. Kahneman, D. & Tversky, A. On the interpretation of intuitive probability: A reply to Jonathan Cohen. (1979).

37. Konstantopoulos, S. Fixed effects and variance components estimation in three-level meta-analysis. en. *Research Synthesis Methods* **2.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.35, 61–76. ISSN: 1759-2887. `https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.35` (2021) (2011).

38. Laibson, D. Golden Eggs and Hyperbolic Discounting*. *The Quarterly Journal of Economics* **112,** 443–478. ISSN: 0033-5533. `https://doi.org/10.1162/003355397555253` (2021) (May 1997).

39. O'Donoghue, T. & Rabin, M. Doing it now or later. *American Economic Review* **89.** WOS:000079411900005, 103–124. ISSN: 0002-8282 (Mar. 1999).

40. Pollak, R. A. Consistent Planning.'Review of Economic Studies+-#* : 2010208 (1968).

41. Rabin, M. An approach to incorporating psychology into economics. *American Economic Review* **103,** 617–22 (2013).

42. Samuelson, P. A. A note on measurement of utility. *The review of economic studies* **4,** 155–161 (1937).

43. Tipton, E. Small sample adjustments for robust variance estimation with meta-regression. eng. *Psychological Methods* **20,** 375–393. ISSN: 1939-1463 (Sept. 2015).

44. Van den Noortgate, W., López-López, J. A., Marín-Martínez, F. & Sánchez-Meca, J. Three-level meta-analysis of dependent effect sizes. en. *Behavior Research Methods* **45,** 576–594. ISSN: 1554-3528. `http://link.springer.com/10.3758/s13428-012-0261-6` (2021) (June 2013).

45. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *Journal of statistical software* **36,** 1–48 (2010).

46. Wong, W.-K. How much time-inconsistency is there and does it matter? Evidence on self-awareness, size, and effects. *Journal of Economic Behavior & Organization* **68.** Publisher: Elsevier, 645–656 (2008).

47. Yaouanq, Y. L. & Schwardmann, P. Learning About One's Self (2019).

A Process of the naivety estimate approximation

Let $A = 1 - \hat{\beta}$, $B = 1 - \beta$, and $f(A, B) = \frac{A}{B}$. I Taylor-expand f at the point of mean $\alpha = (\mu_A, \mu_B)$ to the first order.

$$f(A, B) = f(\alpha) + f'_A(A - \mu_A) + f'_B(B - \mu_B) + R$$

where R is the remainder of the expansion. Thus we have:

$$\begin{aligned} E[f(A, B)] &= E[f(\alpha) + f'_A(A - \mu_A) + f'_B(B - \mu_B) + R] \\ &\sim E[f(\alpha) + f'_A(A - \mu_A) + f'_B(B - \mu_B)] \\ &= \frac{\mu_A}{\mu_B} \end{aligned}$$

I then used this result to study the approximate variance:

$$\begin{aligned} Var[f(A,B)] &= E[f(A,B) - E[f(A,B)]]^2 \\ &\sim E[f(A,B) - f(\alpha)]^2 \\ &= \frac{1}{\mu_B^2} Var(A) - 2\frac{\mu_A}{\mu_B^3} Cov(A,B) + \frac{\mu_A^2}{\mu_B^4} Var(B) \\ &= (\frac{\mu_A}{\mu_B})^2 [\frac{\sigma_A^2}{\mu_A^2} + \frac{\sigma_B^2}{\mu_B^2} - 2\frac{Cov(A,B)}{\mu_A \mu_B}] \end{aligned}$$

This tells us that to obtain the approximate variance of $\frac{1-\hat{\beta}}{1-\beta}$, it is essential for us to understand the covariance between each pair of β and $\hat{\beta}$ estimates, as these two estimates are likely correlated with each other.

As we know that $E[AB] = E[A]E[B]$ only if $Cov(A,B) = 0$. And by construction, β and $\hat{\beta}$ are arguably correlated with each other. Thus, we aim for a more accurate approximate for the mean item. Consider a second-order taylor expansion at $\alpha = (\mu_A, \mu_B)$:

$$\begin{aligned} f(A,B) &= f(\alpha) + f'_A(A - \mu_A) + f'_B(B - \mu_B) \\ &\quad + \frac{1}{2}[f''_{AA}(\alpha)(A - A_\alpha)^2 + f''_{BB}(\alpha)(B - B_\alpha)^2 - 2f''_{AB}(\alpha)(A - A_\alpha)(B - B_\alpha)] + R \end{aligned}$$

This in turn implies that:

$$\begin{aligned} E[f(A,B)] &\sim E[f(\alpha)] \\ &\quad + E\{\frac{1}{2}[f''_{AA}(\alpha)(A - A_\alpha)^2 + f''_{BB}(\alpha)(B - B_\alpha)^2 - 2f''_{AB}(\alpha)(A - A_\alpha)(B - B_\alpha)]\} \\ &= \frac{\mu_A}{\mu_B} - \frac{Cov(A,B)}{\mu_B^2} + \frac{\mu_A}{\mu_B^3} Var(B) \end{aligned}$$

The derivation starts from plugging in $f(A,B) = \frac{A}{B}$ and calculating the respective derivatives. From this, we can derive an approximate mean and variance

vector $(\mu_{naivety}, \sigma^2_{naivety})$ for the naivety measure$\frac{1-\hat{\beta}}{1-\beta}$. Then we have:

$$\mu_{naivety} \sim \frac{1-\mu_{\hat{\beta}}}{1-\mu_\beta} - \frac{Cov(\beta,\hat{\beta})}{(1-\mu_\beta)^2} + \frac{1-\mu_{\hat{\beta}}}{(1-\mu_\beta)^3}\sigma^2_\beta$$
$$\sigma^2_{naivety} \sim (\frac{1-\mu_{\hat{\beta}}}{1-\mu_\beta})^2[\frac{\sigma^2_{\hat{\beta}}}{(1-\mu_{\hat{\beta}})^2} + \frac{\sigma^2_\beta}{(1-\mu_\beta)^2} - \frac{2Cov(\beta,\hat{\beta})}{(1-\mu_\beta)(1-\mu_{\hat{\beta}})}]$$

by plugging in $A = 1-\hat{\beta}$, $B = 1-\beta$.

B RVE variance estimation

This appendix section offers a full construction of two estimators $\hat{\tau}^2$ and $\hat{\omega}^2$. The derivation follows the Erratum of Hedges et al (2010). In a RE-RVE model, assume all error terms indepdent to each other. Let n_{ij} denote the jth naivety estimate in the ith study in the data set. Denote variance of estimates in study i as V_i, then by the independence assumption and the form of RE-RVE model, we have:

$$V_i = \tau^2 + \omega^2 + \tilde{v_i}^2$$

with the first two terms on the right hand side defined as in the main article. $\tilde{v_i}^2$ is a measure of estimates' variance in study i, which usually takes the form of average. The optimal weight for each study then is

$$W_i^* = \frac{1}{\tau^2 + \omega^2 + \tilde{v_i}^2}$$

Here, both τ^2 and ω^2 need to be estimated as it is impossible to observe them directly. Following Hedges et al (2010), the two corresponding estimators are:

$$\hat{\tau}^2 = \frac{Q - C_2}{A_2} - \hat{\omega}^2$$
$$\hat{\omega}^2 = \frac{A_2(Q_1 - C_1) - A_1(Q - C_2)}{B_1A_2 - A_1B_2}$$

Firstly, Q here is the Q-statistic calculated as usual. Let $w_i = \frac{1}{\tilde{v}_i}$ be the FE-optimal weight,

$$Q = \sum_i \sum_j w_i n_{ij}^2 - \frac{(\sum_i \sum_j w_i n_{ij})^2}{\sum_i \sum_j w_{ij}}$$

Similar to Q, Q_1 here is another form of variance using the fixed-effect synthesis as the mean. That is,

$$Q_1 = \sum_i \sum_j (n_{ij} - \frac{\sum_i \sum_j w_{ij} n_{ij}}{\sum_i \sum_j w_{ij}})$$

Last, $A_1, A_2, B_1, B_2, C_1, C_2$ are weights associated with both estimators. Let k_i denote the number of estimates within study i:

$$\begin{aligned}
A_1 &= \sum_m k_i^2 - 2\frac{\sum_i k_i w_i}{\sum_i w_i} + \frac{m \sum_i w_i^2}{(\sum_i w_i)^2} \\
A_2 &= \sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i} \\
B_1 &= \sum_m k_i - 2\frac{\sum_i k_i w_i}{\sum_w i} + \frac{m \sum_i w_i^2}{(\sum_i w_i)^2} \\
B_2 &= \sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i} \\
C_1 &= \sum_i \frac{1}{w_i} - m\frac{1}{\sum_i w_i} \\
C_2 &= \sum_i k_i - 1
\end{aligned}$$