Meta-Analysis of Empirical Estimates of Loss Aversion

Alexander L. Brown Taisuke Imai
Ferdinand M. Vieider Colin F. Camerer *

June 2, 2022

Abstract

Loss aversion is one of the most widely used concepts in behavioral economics. We conduct a large-scale, interdisciplinary meta-analysis, to systematically accumulate knowledge from numerous empirical estimates of the loss aversion coefficient reported from 1992 to 2017. We examine 607 empirical estimates of loss aversion from 150 articles in economics, psychology, neuroscience, and several other disciplines. Our analysis indicates that the mean loss aversion coefficient is 1.955 with a 95% probability that the true value falls in the interval $[1.820, 2.105]$. We record several observable characteristics of the study designs. Few characteristics are substantially correlated with differences in the mean estimates.

JEL Classification codes: D81, D90, C90, C11
Keywords: loss aversion, prospect theory, meta-analysis

*Brown: Department of Economics, Texas A&M University, alexbrown@tamu.edu. Imai: Department of Economics, LMU Munich, taisuke.imai@econ.lmu.de. Vieider: RISL$\alpha\beta$, Department of Economics, Ghent University; and RISL$\alpha\beta$ Africa, AIRESS, University Mohammed VI Polytechnic, ferdinand.vieider@ugent.be. Camerer: Division of the Humanities and Social Sciences, California Institute of Technology, camerer@hss.caltech.edu. We thank Hyundam Je and Hye Joon Lee for research assistance. We are grateful for the feedback provided by Stefano DellaVigna, Chishio Furukawa, Tomáš Havránek, Klaus Schmidt, Tom Stanley, and the audience at ESA 2019 World Meeting, ESA North American Meeting 2019, MAER-Net Colloquium 2019, D-TEA 2020, 2020 CESifo Area Conference on Behavioral Economics, M-BEES 2021, NASMES 2021, Barcelona GSE Summer Forum 2021 (External Validity, Generalizability and Replicability of Economic Experiments), SPUDM 2021, EEA-ESEM 2021, VfS Annual Conference 2021, and seminars at Hitotsubashi University, Appalachian State University, Carleton University, and BIBaP. Financial support from the National Science Foundation #1757282 (Brown and Camerer) and Deutsche Forschungsgemeinschaft through CRC TRR 190 (Imai) are gratefully acknowledged. The data and the analysis code are available at the OSF repository: https://osf.io/9un34/.

1 Introduction

Loss aversion is the empirical observation that decisions often reflect a larger distaste for potential losses, compared to equal-sized gains, relative to a point of reference. Loss aversion is a core feature of prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Wakker, 2010), an explicitly descriptive model of choice under risk and uncertainty which has been widely applied and cited.[1] The strength of aversion to loss compared to attraction to gain is typically captured by a single parameter, λ.

In his popular-science book, Kahneman (2011) writes that "the concept of loss aversion is certainly the most significant contribution of psychology to behavioral economics" (p. 300). Loss aversion has been widely applied to many types of economic decisions and analyses. It is often applied in analyses of experimental decisions over monetary risks (as in the original Kahneman and Tversky, 1979). However, the use of loss aversion and dependence on reference points has evolved well beyond its initial application. Applications include financial asset prices (Barberis, 2013), the equity premium puzzle (Benartzi and Thaler, 1995), labor supply decisions (Camerer et al., 1997), political power of entitlements change (Romer, 1996), majority voting and politics (Alesina and Passarelli, 2019), sectoral trade policy behavior (Tovar, 2009), and selling-buying price endowment effects in contingent valuation of nontraded goods (Ericson and Fuster, 2014; Tunçel and Hammitt, 2014). Loss aversion also features prominently in behavioral industrial organization, in theories and evidence of responses to price changes (Heidhues and Kőszegi, 2018).

Several different methods have been used to measure loss aversion. These include laboratory experiments, representative panel surveys, analyses of natural data, and randomized trials trying to change behavior. Loss aversion has been quantified for monetary outcomes as well as for non-monetary outcomes, such as health (Attema, Brouwer and l'Haridon, 2013). Other fields outside economics also utilize loss aversion, including neuroscience, psychiatry, business and management, and transportation.

Given how widely the concept of loss aversion has been applied in economics and many other social sciences, it is useful to have the best possible empirical answer about how large loss aversion is, and how it varies. One of the first empirical estimates of λ is reported in Tversky and Kahneman (1992). The authors elicit the preferences of 25 graduate students from elite west-coast American universities using three sessions of unincentivized lottery-choice experiments. The median λ—no mean nor statistic of dispersion was reported—was $\lambda = 2.25$ (Figure 1). Many analyses, to this day, cite this number as the typical degree of loss aversion. For example, the value $\lambda = 2.25$ is used in numerical simulations of prospect

[1]The 1979 paper is the most widely cited empirical economics paper published from 1970-2005 (see Table 2 in Kim, Morse and Zingales, 2006). Note also that Fishburn and Kochenberger (1979) documented loss aversion in a different sample of preferences elicited for decision analysis.

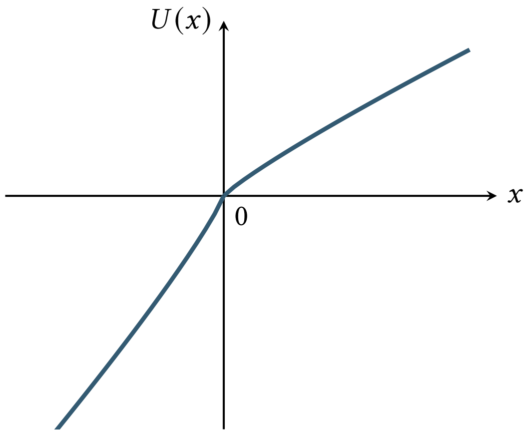

FIGURE 1: An example of prospect theory utility function. *Notes*: This is the specification (3) presented in Section 2, $U(x) = x^\alpha$ for $x \geq 0$ and $U(x) = -\lambda(-x)^\beta$ for $x < 0$, with median parameters of $\lambda = 2.25$ and $\alpha = \beta = 0.88$ reported in Tversky and Kahneman (1992).

theory in behavioral finance (e.g, Barberis, Huang and Santos, 2001; Barberis and Huang, 2001, 2008; Barberis and Xiong, 2009; Barberis, Mukherjee and Wang, 2016; Barberis, Jin and Wang, 2021). Of course, had Tversky and Kahneman initially reported a different value (e.g., 1.5) these analyses might yield different findings. Some of these authors are well-aware of this issue.

As noted in the latter study, "[...] these estimates are almost 30 years old and are based on a small number of participants. Given that the values we assign to these parameters play a significant role in our results, it seems prudent to base these values on a wide range of studies, not just one." (Barberis, Jin and Wang, 2021, p. 2665).[2]

What is the best way to cumulate knowledge about λ after thirty years of research? Our view is that *meta-analysis* is an indispensable tool for scientific cumulation. Meta-analysis is a principled, reproducible, open-science method for accumulating scientific knowledge (and also for detecting nonrandom selective reporting of evidence: Stanley, 2001; Stanley and Doucouliagos, 2012). A meta-analysis uses a clearly specified method of sampling available studies, coding evidence in a way that is comparable across studies, and summarizing both regularity and variation across studies. The idea of synthesizing evidence from multiple studies dates back to the early 1900s (Pearson, 1904; Yates and Cochran, 1938), but the history of

[2]It is not our contention that there is a consensus in the academic community around the estimate $\lambda = 2.25$, though there may be some anchoring to that value (as we have just described). Our general contention is that there is no consensus and lack of confidence in a uniform estimate. In fact, during one of the early presentations of this paper at the Economic Science Association World Meeting in Vancouver in 2019, we elicited guesses of our mean parameter. We incentivized the audience to guess correctly with a CA$50 dollar prize for the closest guess. We have 37 guesses and 34 participants also reported their confidence levels (low, medium, or high). Mean guesses (of the mean parameter) were 1.639 with standard deviation of 0.599. Of the 34 answers, 20 (58.8%) reported low confidence in their guesses and nine (24.3%) fell between 1.8 and 2.1. See Online Appendix F for the full distribution of the guesses.

modern meta-analysis has its origin in the 1976 AERA presidential address by Gene V. Glass. He introduced the term "meta-analysis" to refer to "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass, 1976, p. 3). It has been widely used in evidence-based practices in medicine and policy for at least two decades (Gurevitch et al., 2018). However, meta-analysis has been mostly absent from highly-selective journals in empirical economics.[3]

This paper reports the results of a meta-analysis of empirical estimates of loss aversion. The dataset comprises 607 estimates reported in 150 papers in economics, psychology, neuroscience, and several other disciplines.

The toolkit of meta-analysis can give the best available answers to three questions:

1. What is the central tendency in the distribution of λ estimates; and how much do they vary?

2. Does measured λ vary systematically across different methods, definition of λ, utility specifications, domains of choice, and types of participants?

3. Is there evidence of selective reporting, or publication bias, which distorts reported estimates of λ compared to the corpus of ideal evidence without such biases?

While the answers to these questions no doubt carry some intrinsic interest to researchers interested in loss aversion, they also have practical validity. For one, this paper's mean estimate of $\lambda = 1.955$ provides a much more informed value of loss aversion than the original $\lambda = 2.25$ for researchers to use as an input in financial models (see above). The results can help researchers do their work better in several other ways.

Imagine a researcher who is interested in loss aversion but not quite sure what steps to take to measure it or to apply it. First, the researcher might ask: What method should I use to measure λ? What are the most popular methods? Does it make much difference which one is used? Results on how estimated λ's vary with characteristics of the measurement method, such as the type of the data (experimental or field), reward (monetary or non-monetary), specification of the utility function, and the definition of loss aversion, can guide the researcher.[4]

[3]Prominent meta-analyses in economics include value of a statistical life (Doucouliagos, Stanley and Giles, 2012; Doucouliagos, Stanley and Viscusi, 2014), intertemporal elasticity (Havránek, 2015; Havránek et al., 2015), habit formation (Havránek, Rusnak and Sokolova, 2017), foreign direct investment (Iršová and Havránek, 2013), minimum wage effects (Card and Krueger, 1995; Doucouliagos and Stanley, 2009), gender wage discrimination (Stanley and Jarrell, 1998), microcredit interventions (Meager, 2022), behavior in dictator and ultimatum games (Engel, 2011; Oosterbeek, Sloof and van de Kuilen, 2004), preferences for truth-telling (Abeler, Nosenzo and Raymond, 2019), experimentally-measured discount rates (Matoušek, Havránek and Iršová, 2022), and present-bias in Convex Time Budget experiments (Imai, Rutter and Camerer, 2021).

[4]These methodological variations can also help us understand the mechanisms behind loss aversion, along with process measures such as response times, psychophysiology, and neuroscientific data.

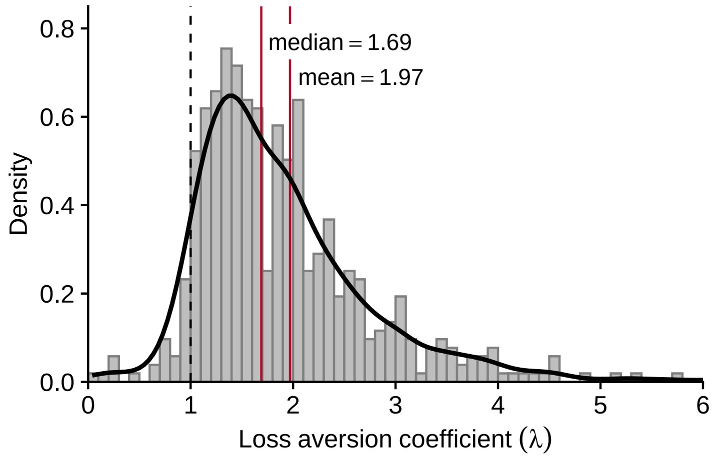

FIGURE 2: Distribution of reported estimates of loss aversion coefficient. *Notes*: Bins for histogram are 0.1 wide. Kernel density estimate of the distribution is superimposed, using the Gaussian kernel with the Silverman's rule of thumb for the bandwidth selection. There are 85 cases that report both individual-level mean and median. We keep individual-level medians from these cases. The x-axis is cut off at 6 for better visual rendering but the density estimation keeps five observations with $\lambda > 6$.

Second, the researcher may be doing a behavior change intervention leveraging the psychology of loss aversion. Then she needs a specific estimate of λ, or a plausible range of values, to use to make a power calculation. Perhaps she is planning to prepay teacher bonuses, which they can later lose, to motivate them to increase student test outcomes (Fryer et al., 2012). Is $\lambda = 2.25$ a good guess or is there a better guess? Is there a more refined estimate of λ for the subset of studies in the meta-analysis which are most like the one she is planning? Meta-analysis can help here too.

Third, suppose the researcher has just read review articles about prospect theory and reference-dependent preferences (e.g., Barberis, 2013; DellaVigna, 2009, 2018; O'Donoghue and Sprenger, 2018). Those reviews have a "narrative" programmatic structure in which results of early and key studies raise fundamental questions that later studies are designed to answer. The reader is usually left with an understanding of the historical intellectual trajectory, and what the next wave of studies should try to understand better. The researcher wonders, is anything important left out of the narrative? The meta-analysis helps answer this question too. (However, the comparison and complementarities of meta-analysis and narrative review are subtle and important, so we will return to them in the conclusion.)

Figure 2 shows the distribution of loss aversion coefficients λ in our dataset, where the median value of the raw data points is 1.69 and the mean is 1.97. The distribution is right-skewed and has a substantial mass (93.9%) on the range $\lambda > 1$, corresponding to loss aversion (as opposed to loss tolerance, $\lambda < 1$). Applying a Bayesian hierarchical approach taking into account the uncertainty surrounding the measurements, we find that the average λ in the literature lies between 1.7 and 1.9. Taking into account the fact that many papers reported more than one estimate (thus producing correlation among estimates), the average is between 1.8 and 2.1. We also examine whether observed heterogeneity in reported λ can be attributed to some of the observable characteristics of the study design. The results do not show many strong reliable effects.

Even to economists unfamiliar with meta-analysis, the method should be, in some ways, familiar. It is essentially an application of econometric techniques to literature review (see Stanley and Doucouliagos, 2012). Like for any empirical study, the greatest concern should be the inclusion criteria of the dataset (i.e., selection). While our broad inclusion criteria are independent of estimates of λ, we have little control over publication decisions (which makes papers more prominent and easier to find) and whether a study is written at all (i.e., "the file drawer problem"; Rosenthal, 1979), which could be dependent on the values of estimated λ. We consider these issues and how they might affect our analysis by examining the correlation between estimated λ and their standard errors and by inspecting the shape of distributions of z-scores. As noted elsewhere, no technique, not even the alternative narrative review, can address this issues perfectly (Borenstein et al., 2009). Meta-analysis at least has the tools to examine these possible issues quantitatively.

Finally, we note an advantage of meta-analysis is that as new evidence arrives it can be easily added to the previous corpus of studies and results can be quickly updated. To allow researchers to achieve this purpose, we provide the most up-to-date materials, the data, the analysis code, and the list of additional articles that were not included in our current meta-analysis (e.g., articles that appeared after our cut-off for inclusion or mentioned by the original authors), on the project repository at the Open Science Framework (`https://osf.io/9un34/`).

Related papers. There are two previous meta-analyses of loss aversion, to which we contribute a newer and broader scope.[5] Neumann and Böckenholt (2014) conducted a meta-analysis of 109 estimates of loss aversion from 33 studies about consumer brand choice. As we do later in this paper, they use a multi-level, random-effects technique to account for variability of estimates, both within and between studies, of the logged-λ parameter. They report a base model estimate of $\lambda = 1.49$ and an "enhanced model" estimate of $\lambda = 1.73$ accounting for sources of estimate variability. Perhaps because of their narrow focus on consumer choice, their meta-regression controls explain nearly all of the variability within their data. Notably, the use of external vs. internal reference points, estimates derived from models that account for both heterogeneity in taste and process, and unpublished vs. published studies are all associated with lower estimates of loss aversion.

[5] Mrkva et al. (2020) ask a question similar to ours, examining how individual differences moderate the degree of loss aversion. Their approach is different— they conduct five large-scale field surveys with a total of 17,720 subjects.

A different approach was used by Walasek, Mullett and Stewart (2018) to understand heterogeneity in a narrow domain of mixed gain-loss financial lotteries. Their analysis used only published experimental studies of mixed lotteries of gains and losses where original raw data were available for reanalysis. Their corpus is 19 estimates from 17 articles.[6] Rather than meta-analyzing estimates from the original papers, they re-estimated parameters for a single model of cumulative prospect theory (i.e., power utility function with symmetric curvature, $\alpha = \beta$, see equation (3) in Section 2) using the original data. Their random-effects meta-analysis on the 19 estimates has an average $\lambda = 1.31$. Despite their rather strict restrictions, the authors note that there are high levels of methodological variability between studies (their data is not very useful in looking at this question within studies) in both estimates and procedures.

The rest of the paper is organized as follows. Section 2 introduces the concept of loss aversion in prospect theory. Section 3 describes how we assembled the dataset of empirical estimates of loss aversion. Section 4 provides results and Section 5 discusses their implications.

2 Loss Aversion

In this section, we briefly illustrate some typical definitions of loss aversion in prospect theory. Consider a situation where an agent makes a choice under risk between prospects with at most two distinct outcomes. This simplified structure still captures a wide range of empirical studies examined here. Let $(x, p; y)$ denote a *simple lottery*, which gives outcome x with probability p and outcome y with probability $1-p$ (Abdellaoui, Bleichrodt and Paraschiv, 2007; Chateauneuf and Wakker, 1999; Köbberling and Wakker, 2005). A key assumption of prospect theory is that outcomes are evaluated as *gains* and *losses* relative to a *reference point*. For simplicity of exposition, in this section, we assume the reference point to be 0, so that the sign of the outcome indicates whether it is a gain or a loss. We call a lottery *non-mixed* if two outcomes have the same sign (i.e., either $x, y \geq 0$ or $x, y \leq 0$) and *mixed* if one of the outcomes is positive and the other outcome is negative. Without loss of generality, we assume that $x > 0 > y$ when we deal with a mixed lottery.

In this setup, both original prospect theory by Kahneman and Tversky (1979) (hereafter OPT) and its modern incarnation, cumulative prospect theory of Tversky and Kahneman (1992) (hereafter PT), postulate that the agent evaluates non-mixed prospects $(x, p; y)$ with

[6] Though not specifically excluded by the aforementioned criteria, the authors also excluded studies that relied on adaptive questions because of concerns about how such techniques would affect their maximum likelihood estimation procedures (i.e., Abdellaoui, Bleichrodt and l'Haridon, 2008; Wakker and Deneffe, 1996).

$x \geq y \geq 0$ or $x \leq y \leq 0$ by

$$w^s(p)U(x) + (1 - w^s(p))\,U(y), \tag{1}$$

and mixed prospects $(x, p; y)$ with $x > 0 > y$ by

$$w^+(p)U(x) + w^-(1-p)U(y), \tag{2}$$

where $w^s : [0,1] \to [0,1]$ is a probability weighting function for gains ($s = +$) or for losses ($s = -$), with $w^s(0) = 0$ and $w^s(1) = 1$, and $U : \mathbf{R} \to \mathbf{R}$ is a strictly increasing utility function satisfying $U(0) = 0$. Tversky and Kahneman (1992) assume that the utility function U and the probability weighting functions w^+ and w^- exhibit diminishing sensitivity. Note also that $w^+ = w^-$ is assumed under OPT, and expected utility is a special case of PT where $w^+(p) = w^-(p) = p$ for all $p \in [0,1]$.

A particularly popular functional apparatus is the one using different power utility parameters for gains and losses, following the approach of Tversky and Kahneman (1992):

$$U(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -\lambda(-x)^\beta & \text{if } x < 0 \end{cases} \tag{3}$$

where $\lambda > 0$ is the *loss aversion coefficient*, the target variable of interest in this study. Values of $\lambda > 1$ are taken to indicate loss aversion, whereas values of $\lambda < 1$ indicate loss tolerance (also referred to as "gain seeking"), with $\lambda = 1$ indicating loss neutrality. The utility function is concave for gains and convex for losses, reflecting diminishing sensitivity, when $\alpha, \beta \in (0,1)$.

Notice that mixed prospects are necessary to identify loss aversion, since λ cancels out in the evaluation of pure-loss prospects such as in equation (1). In this particular formulation, the loss aversion parameter is dependent on the scale of the data, and thus not uniquely defined due to scaling issues (see Wakker, 2010, Section 9.6, for a theoretical discussion). If, on the other hand, the two power parameters are assumed to be identical, i.e. $\alpha = \beta$, this issue does not occur. It also does not occur for different utility parameters using alternative functional forms, such as exponential utility (Köbberling and Wakker, 2005).

Beyond the popularity of the formulation provided above, it is important to note that several different definitions have been proposed and used in the literature. Köbberling and Wakker (2005) and Abdellaoui, Bleichrodt and Paraschiv (2007) provide extensive discussions of such alternative definitions, which we summarize in Online Appendix B. Furthermore, under PT decision weights (given by probability weighting functions w^+ and w^-) naturally enter the definition of loss aversion (Schmidt and Zank, 2005). The combination of different definitions with different functional forms for utility and weighting functions results in a large variety of different formulations. Our strategy in this meta-analysis is simply to take the estimate emerging from the formulation of the authors. We code the type of definition adopted, to be able to determine the correlation of definitions and functional forms with estimates.

A definition that more clearly departs from the apparatus presented above is the expectation-based reference-dependent model of Kőszegi and Rabin (2006, 2007). In this model, an agent evaluates a consumption outcome x by

$$v(x \mid r) = m(x) + \mu\big(m(x) - m(r)\big),$$

where the function m represents the direct utility from consumption and the function μ represents the "gain-loss" utility from departures from a reference point r. In typical applications of the model the consumption utility m is assumed to be linear, so that $m(x) = x$, and a piecewise-linear gain-loss utility function is adopted:

$$\mu(z) = \begin{cases} \eta z & \text{if } z \geq 0 \\ \lambda \eta z & \text{if } z < 0 \end{cases}$$

where the parameter $\eta \geq 0$ captures the importance of the gain-loss utility relative to the consumption utility, and λ again captures loss aversion.[7]

3 Data

3.1 Identification and Selection of Relevant Studies

In order to deliver an unbiased meta-analysis, we first identified and selected relevant papers following unambiguously specified inclusion criteria. The main criterion is to include "all empirical papers that estimate a coefficient of loss aversion." Note that, under this criterion, we include papers that use choice data from laboratory or field experiments and also non-experimental, naturally occurring data including stock prices, TV game shows, and surveys on transportation.

We searched for relevant papers on the scientific citation indexing database Web of Science. The initial search, made in the summer of 2017, returned a total hits of 1,547 papers. As a first step of paper identification, we went through titles and abstracts and threw out 910 papers that were clearly irrelevant for our study. We then read the remaining papers, applied our inclusion criteria based on the content, and then coded information (described in Section 3.2 below). We also used IDEAS/RePEc and Google Scholar to search for unpublished working

[7] For small to modest-scale risks, consumption utility can be taken to be approximately linear (Kőszegi and Rabin, 2007; Rabin, 2000).

papers. Finally, we posted a message on the email list of the Economic Science Association to ask for relevant papers (in February 2018).

In the initial search phase, we cast a net also to identify papers that estimate the degree of loss aversion in riskless choices, through measuring the discrepancy between willingness-to-pay (WTP) and willingness-to-accept (WTA). However, in reading many of these papers, we discovered that while the authors had measured WTP and WTA for a given item, they had not intended to do so as a way to estimate the loss aversion coefficient. While it is straightforward to impose a certain linear utility structure on such papers to recover a loss aversion coefficient, we viewed it as problematic to impose our own assumptions on the work of others, and thus, faced with no better option, did not include these papers.

The search and selection procedure is summarized in Online Appendix A.1. We identified 150 papers at the end of this process. Twenty papers are unpublished at the time of the initial data collection (summer 2017).

3.2 Data Construction

We assembled the dataset for our meta-analysis by coding relevant information—estimates of the loss aversion coefficient and the associated standard error, characteristics of the data, and measurement methods. The primary variables of interest are estimates of the loss aversion coefficient λ. These estimates come in two different forms: (i) *aggregate-level*, where a single λ for the "representative" agent/subject is estimated by pooling data from all subjects in a study; (ii) *individual-level*, where λ is estimated for each subject in a study and the summary statistics of empirical distribution, typically mean or median, are reported. We have a dummy variable capturing the type of reported estimates. We also coded standard errors (SEs) of parameter estimates as a measure of the estimate's uncertainty/precision and the study's quality. All conventional meta-analyses require SEs to calculate weighted averages and to correct for the heteroskedasticity of meta-regression. Other measures of study quality, when known, can be easily included in these weights. When SEs are not reported, we reconstructed them from other available information such as standard deviation (SD), p-value (of the null hypothesis of loss neutrality), or the inter-quartile range (IQR).[8]

We also coded variables describing characteristics of the data and measurement methods. These variables include: type of the data (e.g., experimental, non-experimental, TV game show); location of the experiment (e.g., laboratory, field, online); types of reward (e.g., real or

[8] We calculated/approximated 68 SEs from other available information: 64 from IQR and sample size and four from p-values. Use of the IQR to infer the SE of the mean (from an approximation of the standard deviation, SD $\approx 1.35\times$IQR) is in principle only legitimate if the parameters are normally distributed in the population. That assumption is clearly a stretch. Nevertheless, obtaining even an "approximate" SE seemed preferable to dropping the observation entirely, or to making other, even stronger, assumptions allowing us to keep the observation.

hypothetical, money, health, time); subject population (e.g., children, college students, general population, farmers); definition of loss aversion coefficient (as described in Online Appendix B); utility specification (e.g., CRRA, CARA); and several others. Table A.1 in the Online Appendix lists all variables coded in the study.

The set of papers we included spans a wide range of disciplines (see Table A.3 in the Online Appendix). Since fields/journals have different reporting cultures and standards, we could not always retrieve all the necessary information from reading papers. We thus emailed the authors of the papers when some essential summary statistics of the loss aversion coefficient or sample size information were missing.[9]

While study quality is an ongoing concern for meta-analysis in general (and one that we will revisit in our concluding section), we record quantitative values, such as each estimate's precision and the impact factor of the journal in which the study was published, as measures of study quality in line with past precedent (Stanley and Doucouliagos, 2012). Meta-analyses in economics routinely code several dimensions of research quality such as whether the study accounts for endogeneity, is experimental, uses panel data, etc. These factors are then included in a meta-regression to estimate the effects of study quality differences and to isolate the findings from higher quality studies. Here, we recognize quality differences by coding the impact factor of the journal where the study is published, the experimental nature of the study, the subject pool, and the type of reward used.

3.3 Descriptive Statistics

We identified 150 articles that report an estimate of the loss aversion coefficient λ. See Online Appendix G for the full list of articles included in our meta-analysis. Among these, 130 articles were published in 78 journal outlets (including eight articles published in the "Top 5" journals in economics). The dataset includes papers from a variety of disciplines: economics, management, psychology, neuroscience, medicine, psychiatry, agriculture, environment, transportation, and operations research (see the list of journals and their classifications in Table A.2 in the Online Appendix).

We also identified where the data (either experimental or survey) were collected for 147 articles in the dataset. Most of these articles report estimates from data collected in a single country. Ten of them collected data from two to three countries/regions, and three of them (l'Haridon and Vieider, 2019; Rieger, Wang and Hens, 2017; Wang, Rieger and Hens, 2017) conducted large-scale cross-country studies, collecting data from more than 30 countries/regions.

[9]We contacted the authors of 51 papers asking for additional clarification on 175 estimates (i.e., mean, SE, or the number of observations). If the authors did not respond to our initial request, we sent an additional email. Overall, we received 39 responses. Of those, 28 responses were ultimately useful and we recovered additional information on 78 estimates. The remainder had to be imputed.

Table 1: Study characteristics.

	Freq.	%		Freq.	%
Total number of studies	185	100.0			
Data type			*Reward type*		
Lab experiment	98	53.0	Money	154	83.2
Field experiment	29	15.7	Other	14	7.6
Other field data	20	10.8	Consumption good	8	4.3
Classroom experiment	18	9.7	Mixed	5	2.7
Online experiment	17	9.2	Health	2	1.1
Game show	3	1.6	Food	1	0.5
Subject population			Environment	1	0.5
University population	91	49.2	*Continent*		
General	63	34.1	Europe	78	42.2
Farmer	13	7.0	North America	56	30.3
Mixed	5	2.7	Asia	25	13.5
Children	4	2.2	Africa	8	4.3
Elderly	3	1.6	Oceania	7	3.8
Capuchin monkey	1	0.5	South America	6	3.2
Unknown	5	2.7	Multiple	3	1.6

Notes: In two studies, the geographic location of the data is unknown. These studies were run online through Amazon's Mechanical Turk or a mobile app, and the authors did not specify what geographic controls were used.

In total, the estimates of the loss aversion coefficient comprised in our dataset come from 71 countries/regions (see Figure D.1 in the Online Appendix).

Next we look at the basic design characteristics of the studies. We have 185 "studies" reported in 150 papers, where a study is defined by a combination of several variables: type of the data, location of data collection, subject pool, type of reward, and continent of data collection. The frequency of each design characteristic is shown in Table 1.

The majority of our data comes from laboratory experiments, but we also have studies using non-experimental data such as surveys, stock market data, and game shows. Subjects were mostly recruited from the pool of university students or the general population. There is also a small set of studies which recruited special populations such as financial professionals, entrepreneurs, managers, and patients with psychiatric disorders or gambling problems. The type of reward used in the studies is mostly monetary. About three-quarters of the studies were conducted in Europe or North America.

Next, we look at the main variable of interest, the estimated coefficient of loss aversion. We have a total of 607 estimates in the dataset (Table 2). About half of these estimate the degree of loss aversion of a "representative" subject by pooling data from all subjects together (we call these aggregate-level, or simply aggregate, estimates). The other half estimated the

Table 2: Types of estimates.

	All estimates		With SE	
	Freq.	Prop.	Freq.	Prop.
Aggregate-level	281	0.463	220	0.530
Individual-level mean	160	0.264	126	0.304
Individual-level median	166	0.273	69	0.166
Total	607	1.000	415	1.000

Notes: There are 85 cases where both mean and median of the distribution of individual-level estimates are reported. "With SE" indicates the observations where SEs are available. In addition, there are four aggregate-level estimates for which SEs are approximated with reported p-values, and 64 individual-level medians for which SEs are approximated with IQR. SEs are imputed for the rest of 124 observations (see Section 4.2).

coefficient for each individual subject in the study and reported summary statistics of the distribution, either mean or median. There are 85 cases where we have both the mean and the median of the distribution of the loss aversion coefficients estimated at the individual level.

Finally, we look at the specification of the functional form of U and the definition of loss aversion λ (Table 3). There are 302 observations which assume the CRRA form for the utility functions as in equation (3), following Tversky and Kahneman (1992), but 221 of them assume and estimate a common curvature for gains and losses ($\alpha = \beta$). We observe less variation in the specification of reference points and the definition of loss aversion coefficients. Three-quarters of the observations set the reference point at zero, but our dataset also includes studies where reference points are assumed to be subjects' status quo or expectations. More than 80% of the observations estimate the loss aversion coefficient λ as Tversky and Kahneman (1992) define it.

4 Results

We structure the results into three distinct parts. We start from a non-parametric analysis of the reported loss aversion coefficients and their SEs. We subsequently fit random-effects meta-analytic distributions to the data, and document the estimated mean loss aversion. Finally, we conduct a series of meta-regressions to see to what extent we can explain the estimated between-study variance.

Table 3: Utility function, loss aversion, and reference point.

	Freq.	%		Freq.	%
Total number of estimates	522	100.0			
Loss aversion λ			*Functional form of U*		
Tversky-Kahneman	445	85.2	CRRA	302	57.9
Köbberling-Wakker	37	7.1	CARA	53	10.2
Kőszegi-Rabin	9	1.7	Linear	73	14.0
Other	3	0.6	Other parametric	32	6.1
Not reported	28	5.4	Nonparametric	16	3.1
Reference point			Not reported	46	8.8
Zero	394	75.5			
Status quo	60	11.5			
Expectation	18	3.4			
Other / Not reported	50	9.6			

Notes: There are 85 cases where both mean and median of the distribution of individual-level estimates are reported. We keep only one measure from each of these observations. See Online Appendix B for definitions of loss aversion.

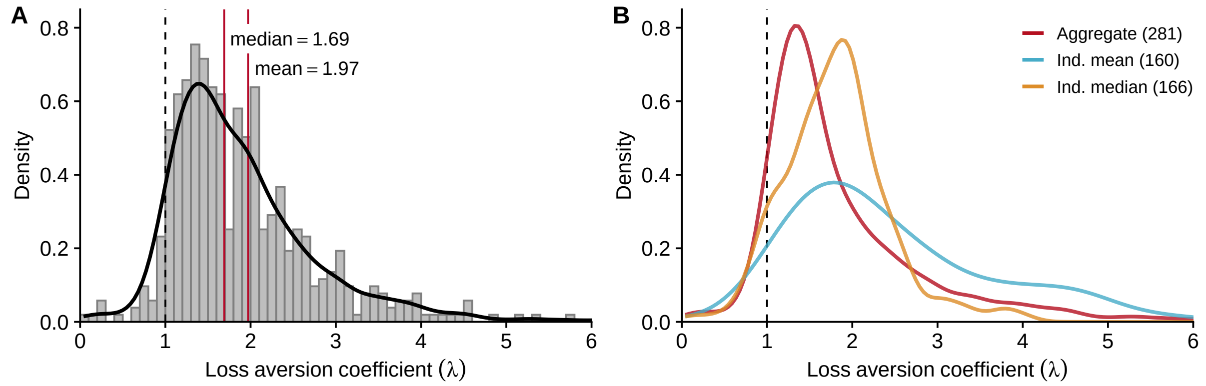

Figure 3: Distribution of loss aversion parameters. (A) All reported estimates combined. (B) Separated by the type of estimates. *Notes*: Panel A is identical to Figure 2. In Panel B, Kernel density estimate of the distribution of reported λ is plotted, using the Gaussian kernel with the Silverman's rule of thumb for the bandwidth selection. All 607 estimates in the data are used for estimation. In the Online Appendix, Figure D.3 shows density plots of $\log(\lambda)$, Figure D.4 shows histograms of λ separated by the type of estimates, and Figure D.5 shows the empirical CDF of λ for each type of estimates.

4.1 Nonparametric Analysis

We start our presentation of the results by showing some non-parametric patterns in the reported loss aversion coefficient λ. Since we do not need SEs for this analysis, we can make use of all the estimates of loss aversion we coded in the dataset. Figure 3A shows the distribution of all the coded loss aversion parameters. The mean of the parameters is 1.97. Given the right skew in the distribution, the median is considerably lower than the mean, at 1.69.

Table 4: Summary statistics of reported λ.

Type	n	Mean	SD	Q1	Median	Q3	Min	Max
Aggregate-level	281	1.950	1.681	1.274	1.520	2.180	0.040	23.460
Individual-level mean	160	2.935	2.605	1.618	2.180	3.395	0.110	19.861
Individual-level median	166	1.844	0.756	1.417	1.800	2.090	0.110	7.500
All	522	1.970	1.370	1.310	1.690	2.288	0.040	23.460

Notes: There are 85 cases that report both individual-level mean and median. We keep individual-level medians from these cases in the last row of the table.

Figure 3B shows the same set of estimates, but now plots separate density functions for the estimates obtained from aggregate-level means ($n = 281$), individual-level means ($n = 160$), and medians ($n = 166$). Aggregate-level estimates can be seen to have the lowest mode (around 1.33), with individual-level medians having a slightly higher mode around 1.88. Means of individual-level estimates show a fat right tail, indicating a higher frequency of larger values. Table 4 shows summary statistics of reported λ for each type of measurement. The means of aggregate-level estimates and individual-level medians are close together, at 1.95 and 1.84, respectively. The somewhat lower mean of the latter results from fewer very large observations amongst medians than amongst aggregate estimates. The individual-level means have the largest variation (SD = 2.61 versus 1.68 for aggregate means and 0.76 for individual-level medians), including some of the smallest estimates as well as some of the largest. The truncated nature of the distribution then results in the highest mean by far, 2.94.

The differences between measurement types above cannot be interpreted causally. That is, the different measurements generally derive from different studies and are based on different data, so that the observed differences cannot be directly ascribed to the type of measurement used. To gain insight into the effect of measurement type, we can conduct an analysis based on the 85 studies for which both means and medians are reported. With a mean of the means of 3.47 (median of means, 2.08) and a mean of the medians of 1.71 (median of medians, 1.69), the results confirm the ones for the overall sample (see Figure D.6 in the Online Appendix). That is, the individual-level estimates tend to be rightward skewed, and this strongly affects the aggregate estimate reported in a paper when means of the individual-level estimates are used instead of medians. This issue, however, will be at least partly remedied by the observation that means of individual-level estimates also tend to come with increased SEs, which will in turn lead to increased pooling of the larger estimates in our meta-analytic estimations.

The earliest evidence recorded in our dataset is Tversky and Kahneman's (1992) famous 2.25. See Figure D.2 in the Online Appendix for the time trend of reported estimates. Half of the estimates in the data are found in papers published after 2015. Individual-level estimates appear in the dataset after 2006, in part due to the rise of common experimental elicitation

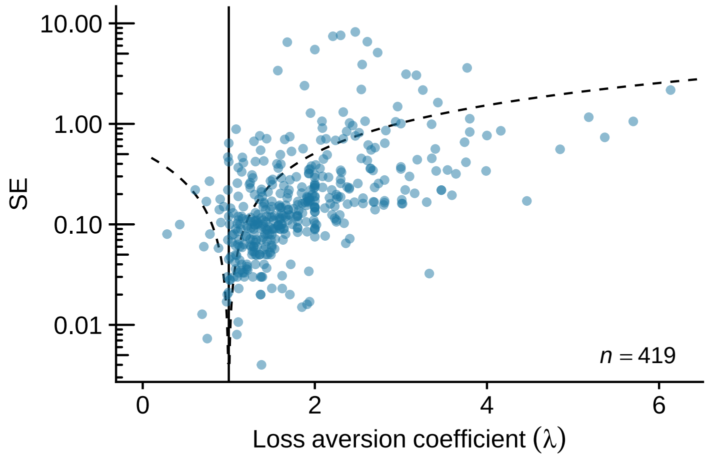

Figure 4: Relationship between reported λ and associated SE. Estimates with corresponding SEs reported are included ($n = 419$). *Notes*: For observations which have both individual-level mean and median, we keep the median in this figure. The vertical solid line corresponds to loss neutrality $\lambda = 1$. Two dashed curves represent the boundaries for statistically-significant loss aversion ($\lambda > 1$) and loss tolerance ($\lambda < 1$). The x-axis is cut off at 6 and the y-axis is displayed in the log-scale for better visualization.

procedures. The raw data do not reveal a clear time trend, suggesting that estimates have remained the same on average for the last 30 years.

Just looking at the raw reported λ, Figure 3 and Table 4 suggest that the "average" loss aversion coefficient λ would locate somewhere between 1.8 and 2.9. At the same time, there is high dispersion in the reported λ. These rough initial estimates, however, do not take the quality of the estimate into account. The latter can be assessed by means of the standard error associated to each estimate, which we can use to calculate a proper "meta-analytic average" by weighing each estimate by its precision, i.e., the inverse of its standard error. Estimates falling far from the mean will then be given less weight to the extent that they have large associated standard errors.

4.2 Precision of Estimates

A common tool of meta-analysis is the so-called "funnel plot" which visualizes the relationship between reported estimates and their associated SEs (Stanley and Doucouliagos, 2012). Figure 4 illustrates this relationship using data points where we have both a value for λ and an associated SE (either reported in the paper or re-constructed from other available information by us).

Figure 4 tells us several things about empirical estimates of loss aversion. Loss aversion may be imprecisely estimated, for instance, because of a small sample size, or because of other characteristics of the design. In our data, estimates that are far from 1 have higher SEs. They are not precisely estimated coefficients due to, for example, small sample size or other issues with the design or behavior. More precisely estimated λ's tend to cluster between 1 and 2. Second, 398 out of 419 estimates (95%) are larger than 1, producing a massive asymmetry in the funnel plot. Third, about 76.6% (305 out of 398) of $\lambda \geq 1$ estimates report results that are significantly different from 1 (based on the two-sided Wald test with the usual significance threshold of $p < 0.05$).

Table 2 above shows that 192 out of our 607 recorded estimates are missing SEs. We approximate SEs from IQRs and p-values for 68 of these observations (see footnote 8), which leave 124 SEs missing. Since standard errors are a fundamental ingredient for meta-analysis because they provide weights for the observations, we thus risk losing many observations, including the iconic measure of 2.25 reported by Tversky and Kahneman (1992). If studies not reporting any SEs are different from studies reporting them, we may furthermore distort our estimates systematically.

To overcome this issue, we impute the missing SEs using the subset of the data for which we have both λ and its associated SE. The basic idea is to estimate the parameters characterizing the distribution of SEs in the data, assuming that $\log(se)$ is drawn from a normal distribution $\mathcal{N}(\mu_{se}, \sigma_{se}^2)$ and that the mean depends on the size of reported λ, $\mu_{se} = \alpha_{se} + \beta_{se}\lambda$. We then impute missing SEs using the estimated distributional parameters $(\widehat{\alpha}_{se}, \widehat{\beta}_{se}, \widehat{\sigma}_{se})$ and reported λ. See Online Appendix A.3 for the detail. Online Appendix C.3.2 also shows that our estimates and conclusions of this paper would not particularly change if these imputed values were dropped. We will, from now on, make use of the full set of observations.

4.3 Average Loss Aversion in the Literature

The main goal of our meta-analysis is first to obtain the "best available" estimate of the loss aversion coefficient λ combining the available information in the literature and then to understand the heterogeneity of reported estimates across studies. Both goals can be informed by the data using a *Bayesian hierarchical modeling* approach.

Setup. Consider the dataset $(\lambda_i, se_i)_{i=1}^m$, where λ_i is the ith *measurement* (or *observation*) of the loss aversion coefficient in the dataset and se_i is the associated standard error that captures the uncertainty surrounding the estimate. In the benchmark model, we assume that the ith reported estimate λ_i is normally distributed around the parameter $\overline{\lambda}_i$:

$$\lambda_i \mid \overline{\lambda}_i, se_i \sim \mathcal{N}(\overline{\lambda}_i, se_i^2), \tag{4}$$

where the variability is due to the sampling variation captured by the known standard error se_i. The parameter $\overline{\lambda}_i$ is often referred to as the "true effect size" in meta-analysis.

Sampling variation is part of the observed variation in the reported estimates $(\lambda_i)_{i=1}^m$, but in addition, there may be "genuine" heterogeneity across measurements (due to different settings, for example). We model this by assuming that each $\overline{\lambda}_i$ is in turn normally distributed, adding another level to the hierarchy:

$$\overline{\lambda}_i \mid \lambda_0, \tau \sim \mathcal{N}(\lambda_0, \tau^2), \tag{5}$$

where λ_0 is the *overall mean* of the estimated loss aversion parameters $\overline{\lambda}_i$, and τ is its standard deviation, capturing the variation between observations in the data. The overall variance in the data, therefore, consists of two parts, the between-observation variance, τ^2, and the individual sampling variation coming from measurement uncertainty, *se*. This can be clearly seen by combining expressions (4) and (5) into one:

$$\lambda_i \mid \lambda_0, \tau, se_i \sim \mathcal{N}(\lambda_0, \tau^2 + se_i^2).$$

Model estimation. We start from fitting the model expressed as equations (4) and (5), restated as model M1 here, to the data $(\lambda_i, se_i)_{i=1}^m$:

$$\begin{aligned} \lambda_i \mid \overline{\lambda}_i, se_i &\sim \mathcal{N}(\overline{\lambda}_i, se_i^2), \\ \overline{\lambda}_i \mid \lambda_0, \tau &\sim \mathcal{N}(\lambda_0, \tau^2), \\ \lambda_0 &\sim \text{half } \mathcal{N}(1, 5), \\ \tau &\sim \text{half } \mathcal{N}(0, 5), \end{aligned} \tag{M1}$$

where "half $\mathcal{N}$" indicates the half-normal distribution which "folds" the normal distribution $\mathcal{N}(0, \sigma^2)$ at its mean to have nonzero probability density for values greater than or equal to 0. This model incorporates the assumption that every observation is statistically independent, and that the observations are normally distributed. We will relax these rather strong assumptions in due time.

We estimate the model in `Stan` (Carpenter et al., 2017) using Hamiltonian Monte Carlo simulations, and launch it from `R` (R Core Team, 2020) using `RStan` (Stan Development Team, 2020). We chose half-normal distributions with a standard deviation of 5 for priors for the population-level parameters λ_0 and τ, but the estimates are not sensitive to changing the prior, given the amount of data we have (see Online Appendix C.3.3).

The estimated overall mean λ_0 is 1.809 with a 95% credible interval (CrI) of [1.740, 1.878].[10,11]

[10] A Bayesian credible interval (CrI) of size $1-\alpha$ given data D is an interval $[L(D), U(D)]$ such that $P(L(D) \leq \theta \leq U(D)) = 1 - \alpha$, where θ is the parameter of interest. Unlike the (frequentist) confidence interval, CrI has a literal probabilistic interpretation: given the data, there is a $100 \times (1-\alpha)\%$ probability that the true parameter value is in the interval.

[11] Results from the frequentist random-effects meta-analysis are presented in Online Appendix E. We obtain

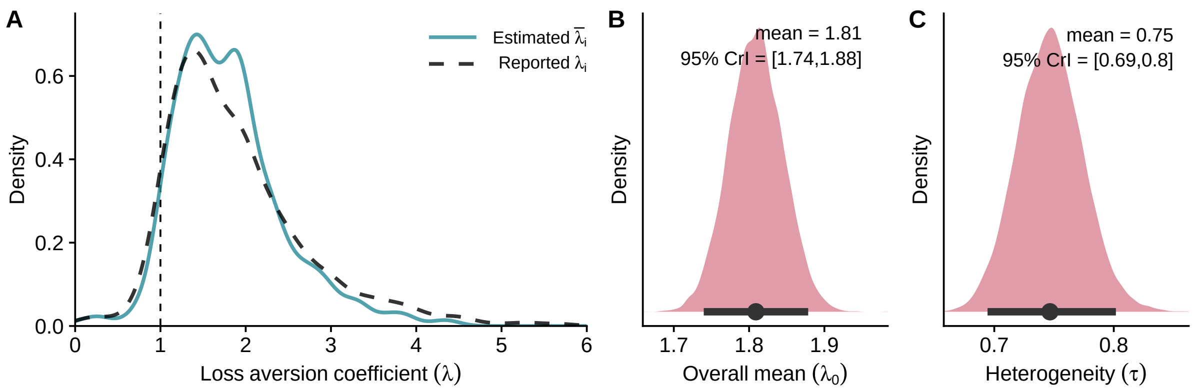

FIGURE 5: (A) Density plot of loss aversion coefficient, estimated $\overline{\lambda}_i$ versus reported λ_i. (B) Posterior draws of the overall mean λ_0. (C) Posterior draws of the heterogeneity parameter τ. *Notes:* The dashed curve in panel A indicates the density of the observed loss aversion parameters, λ_i; the solid curve in panel A indicates the density of the estimated parameters, $\overline{\lambda}_i$. Observations above six are not shown in panel A for better visualization. The black dots and lines in panels B and C represent the posterior means and the 95% credible intervals of λ_0 and τ, respectively.

The mean is clearly lower than the non-parametric result that we saw above, which was 1.97. This is shown in Figure 5AB. The density of estimated $\overline{\lambda}_i$ is lower than the one of observed λ_i for values above 2.5. The same occurs for values below one. This is meta-analytic pooling at work—estimates that fall far from the mean are shrunk towards more plausible values, with the amount of shrinkage proportional to the standard error. See discussion in Online Appendix C.2.

The estimates produced are of course only valid conditional on our assumptions. We already know that the normality assumption seems a stretch, given the skewed distribution of the reported λ. To see this, we can take a look at the *posterior predictive distribution*—the distribution of loss aversion coefficients we would expect new observations λ_{new} to display, provided that the characteristics of the studies from which these observations are obtained are similar on average to those of past studies—and compare it to the distribution of actual observations.[12] This is shown in Figure 6A. Relatively to the actual observations—either as reported (λ_i), or as estimated ($\overline{\lambda}_i$)—the posterior predictive distribution overestimates the likelihood of values smaller than 1, while it underestimates the likelihood of intermediate values between 1 and 2. It does not attribute any probability to values beyond 4, which are not un-

largely identical estimates.

[12]Formally, the posterior predictive distribution is written:

$$\pi\left(\lambda_{\text{new}} \mid (\lambda_i)_{i=1}^m\right) = \int \pi\left(\lambda_{\text{new}} \mid \theta\right) \pi\left(\theta \mid (\lambda_i)_{i=1}^m\right) d\theta,$$

where $\theta = ((\overline{\lambda}_i)_{i=1}^m, \lambda_0, \tau)$ is a vector of model parameters (Gelman et al., 2014). Evaluating this integral is difficult, but we approximate it by drawing $\lambda_{\text{new}}^{(s)} \sim \mathcal{N}(\lambda_0^{(s)}, \tau_{(s)}^2)$ using posterior simulations $(\lambda_0^{(s)}, \tau_{(s)})$, $s = 1, \ldots, N$. We have 8,000 draws (2,000 iterations × 4 chains) of λ_{new}.

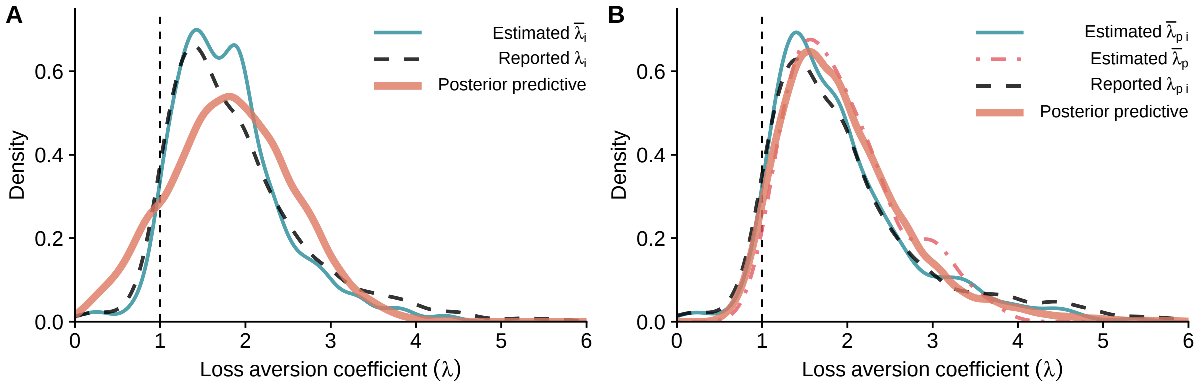

FIGURE 6: Distributions of reported and estimated λ, and posterior predictive distribution of λ. (A) Assuming a normal distribution for the population level (model M1). (B) Assuming a log-normal distribution for the population level (model M2).

common in the data. It thus seems desirable to look for a model that may provide a better fit to the data.

We thus extend the baseline model M1 in two ways. First, we use a log-normal distribution for the population-level distribution. Second, we explicitly model the nesting of observations in papers, in order to overcome any potential distortions deriving from non-independence of observations. Remember that our 607 observations have been obtained from 150 distinct papers, the largest number of observations in a single paper being 53 (Rieger, Wang and Hens, 2017; Wang, Rieger and Hens, 2017). The independence assumption seems rather heroic in this case. To do this, we introduce paper-level estimates as an additional hierarchical level. Let λ_{pi} be the ith estimate reported in paper p. We formulate a model as follows:

$$\begin{aligned}
\lambda_{pi} \mid \overline{\lambda}_{pi}, se_{pi} &\sim \mathcal{N}(\overline{\lambda}_{pi}, se_{pi}^2),\\
\overline{\lambda}_{pi} \mid df, \overline{\lambda}_p, \sigma_p &\sim t(df, \overline{\lambda}_p, \sigma_p^2),\\
\overline{\lambda}_p \mid \lambda_0^\ell, \tau_\ell &\sim \log \mathcal{N}(\lambda_0^\ell, \tau_\ell^2),\\
\lambda_0^\ell &\sim \mathcal{N}(1, 5),\\
\tau_\ell &\sim \text{half } \mathcal{N}(0, 5),\\
df &\sim \text{half } \mathcal{N}(0, 5),\\
\sigma_p &\sim \text{half } \mathcal{N}(0, 5).
\end{aligned} \tag{M2}$$

The model M2 now explicitly models the nesting of the estimated observation-level parameters, $\overline{\lambda}_{pi}$, in paper-level estimates, $\overline{\lambda}_p$. The former are modeled as following a robust student-t distribution instead of a normal distribution to account for observed outliers.[13] The

[13]We estimate the degrees of freedom of the distribution, df, endogenously from the data. This allows us to determine whether the student-t distribution provides a good fit, which is the case if the degrees of freedom are small, or whether it converges to a normal distribution, which is the case for large degrees of freedom (Kruschke,

TABLE 5: Summary of estimation results.

	Distributional assumption			Posterior of λ_0				Posterior of τ			
Model	Obs. level	Paper level	Pop. level	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
M1	Normal		Normal	1.809	0.036	1.740	1.878	0.747	0.027	0.694	0.802
M2	Normal	Student-t	Log-normal	1.955	0.073	1.820	2.105	0.744	0.344	0.603	0.913

Notes: In Model M2, (λ_0, τ) are calculated from the log-normal parameters $(\lambda_0^\ell, \tau_\ell)$ by $\lambda_0 = \exp(\lambda_0^\ell + \tau_\ell^2/2)$ and $\tau^2 = [\exp(\tau_\ell^2) - 1]\exp(2\lambda_0^\ell + \tau_\ell^2)$.

latter are modeled as following a log-normal distribution. Note the super-/sub-scripts ℓ in the location and scale parameters $(\lambda_0^\ell, \tau_\ell^2)$ of the log-normal distribution. We can calculate the mean and the median of the distribution by $\exp(\lambda_0^\ell + \tau_\ell^2/2)$ and $\exp(\lambda_0^\ell)$, respectively, exploiting the properties of the log-normal distribution.

We again start by examining the fit of the model to the data, and by summarizing the population-level parameters. The model fit is shown in Figure 6B. The log-normal distribution can now be seen to fit the estimated paper-level data well. The distribution of the paper-level observations has more probability mass between about 1 and 3, but less beyond that point, compared to the actual study-level observations. The degrees of freedom of the student-t distribution are estimated at 1.32, thus vindicating the use of the robust distribution. The mean loss aversion parameter obtained from this estimation is 1.955, with a 95% CrI of [1.820, 2.105]. Notice, however, that even though this estimate is nearly identical to the one obtained under the standard model at the outset, that occurs by coincidence rather than being a feature of the model. One can further see that there is now increased uncertainty surrounding the prediction interval. This is indeed natural, since the paper-level estimates are surrounded themselves by larger amounts of uncertainty, which is then passed up the hierarchy to the aggregate parameters.

Robustness checks. Online Appendix C.2 presents estimation results for two additional models, but our preferred model M2 fits better than these "intermediate" models. We also estimated the models under different priors or using the "complete" data including only observations where associated SEs are available, and obtained similar conclusions. These robustness checks are presented in Online Appendix C.3.

Heterogeneity between studies versus between individuals. An interesting question concerns how the heterogeneity between studies we document compares to typical levels of heterogeneity between individuals. While our data are ill-suited to answer this question in general, Figure 7 provides an indication by comparing the study-level estimates of loss

2010).

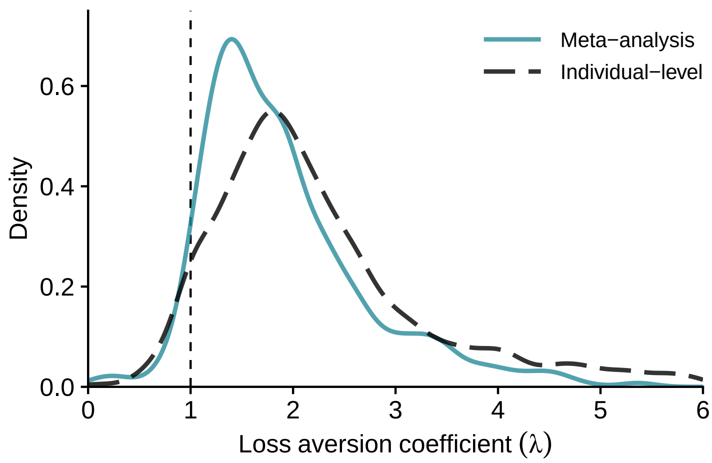

FIGURE 7: Distributions of estimated $\overline{\lambda}_{pi}$, compared to individual-level estimates from l'Haridon and Vieider (2019). Estimates from l'Haridon and Vieider (2019) follow the original hierarchical setup with country-level fixed effects reported in that paper.

aversion obtained from model M2, $\overline{\lambda}_{pi}$, to the between-subject distribution of loss aversion in the data of l'Haridon and Vieider (2019), containing estimates for 3,000 students from 30 countries. Both distributions display a log-normal shape, and have peaks in their densities between 1 and 2. The individual-level estimates are somewhat more dispersed, having slightly more probability mass on small values below 1, as well as more probability mass on values above 2.5. Overall, however, the two distributions are similar. This illustrates just how much heterogeneity we find between studies, which may arguably be driven at least in part by differences in experimental designs, model definitions, and estimation methods.[14]

The comparison shown here comes with a large caveat—individual-level distributions from other papers may look very different. This also goes back to the point on study quality we made above. Given suitable restrictions on choice lists and modeling assumptions, it would be easy to produce individual-level distributions that are narrower than the one shown here. On the other hand, wide choice lists, general definitions, and noisiness in measurements can all contribute to much wider distributions. It can indeed be shown that, given a wide enough range of possible estimates, the location parameter of the log-normal posterior predictive distribution fit to the individual-level estimates will decrease systematically as the proportion of

[14]We note one finding which shows strong between-subject heterogeneity and its association with IQ. Chapman et al. (2018) report data from an incentivized representative survey of Americans measuring loss aversion and other behavioral parameters. In the representative survey, the median λ is 0.99. In a college sample using a highly similar protocol, the median λ is 1.84. In the survey of Americans there are substantial correlations (around 0.2-0.3) between IQ and loss aversion. This paper is not included in our meta-analysis because it fell outside our time window. However, it uses both a different design (optimized adaptive estimation) and features several important within-study measures of heterogeneity, especially IQ. Since this is the only study to report cross-IQ heterogeneity, meta-regression of that feature will add little to general conclusions. Such a study could therefore feature prominently in a narrative review (for reasons discussed in our conclusion).

random choices increases, while the dispersion will increase in random choices.[15] This goes to show just how many different factors may impact the estimation of loss aversion coefficients.

4.4 Explaining Heterogeneity

We observe a non-negligible amount of between-paper heterogeneity (expressed in estimated τ in Table 5, model M2) among reported estimates of λ. In this section, we seek to understand the source of this variability in order to provide a tentative answer to our second key question: "Do reported estimates of λ systematically vary by underlying design characteristics for measurement of loss aversion?"

Remember that we coded several features about the characteristics of study design (Table A.1 in the Online Appendix). Figures D.10 and D.11 in the Online Appendix provide a first look into how these design features are related to reported estimates of λ. Each panel presents how the reported λ varies by underlying design characteristics. We do observe some patterns in the figure, but the effects appear rather weak and it is not clear if these relations are systematic and robust.

We approach this question with a *random-effects meta-regression*, which extends our previous random-effects model by incorporating coded features of the observation or the paper into the model. More precisely, we set up a new model, which expands model M2 by allowing for the location of the observations to be systematically shifted depending on observed characteristics of the observation or the paper. The model looks as follows:

$$\begin{aligned}
\lambda_{pi} \mid \overline{\lambda}_{pi}, se_{pi} &\sim \mathcal{N}(\overline{\lambda}_{pi}, se_{pi}^2),\\
\overline{\lambda}_{pi} \mid df, \overline{\lambda}_p, \sigma_p, \beta &\sim t(df, \overline{\lambda}_p + X_{pi}\beta, \sigma_p^2),\\
\overline{\lambda}_p \mid \lambda_0^\ell, \tau_\ell &\sim \log\mathcal{N}(\lambda_0^\ell, \tau_\ell^2),\\
\lambda_0^\ell &\sim \mathcal{N}(1, 5),\\
\tau_\ell &\sim \text{half }\mathcal{N}(0, 5),\\
df &\sim \text{half }\mathcal{N}(0, 5),\\
\sigma_p &\sim \text{half }\mathcal{N}(0, 5),
\end{aligned}$$

where X_{pi} is a vector of study characteristics associated with ith observation reported in paper p. These characteristics consist mostly of dummy variables taking the value of 0 or 1, with β a vector of coefficients. To facilitate the interpretation of the constant, non-dummy independent variables included in vector X_{pi} are mean-centered—each coefficient in the vector β

[15] Note that the range one can theoretically estimate will depend both on the measurement obtained and the definition of loss aversion. For instance, a narrow choice list in the mixed domain ($x > 0 > y$) may come up with a narrow range of estimations if the definition of loss aversion is taken to be $\lambda = x/(-y)$. For a definition $\lambda = U(x)/(-U(y))$, however, the range resulting from that same choice list could be very wide if the measurements of U allow for extreme values.

then captures the additive effect on the paper-level mean $\overline{\lambda}_p$, relative to the "baseline study" (characterized by the omitted categories in dummy variables and the means of non-dummy variables) which will become clear later.

Estimation results are presented in Figure 8. First, the posterior mean of the estimated $\overline{\lambda}_{p^*}$ for the benchmark study p^* is 1.984 (95% CrI = $[1.593, 2.385]$). Each estimated coefficient in β captures the effect of the study characteristic from this benchmark value.

As we have seen above in Section 4.1, the type of estimates reliably captures the variation in reported λ— individual-level means tend to be higher than the other two types of estimates, due to skewed distributions of individually-estimated λ. We also find that field experiments are associated with higher λ compared to laboratory experiments and studies recruiting general population samples are also associated with higher values of λ compared to the studies with a population of university students. We do not observe differences between studies using monetary rewards and non-monetary rewards, but survey studies tend to produce lower estimates of λ than the binary lottery choice tasks which are common in laboratory experiments. In terms of the specification of the value function, it does not seem to matter much which functional form (CRRA, CARA, etc.) one assumes for the utility functions U, or whether reference points are assumed to be zero, status quo, or expectations. Studies estimating λ following the definition by Köbberling and Wakker (2005) produce higher λ compared to the standard Tversky and Kahneman's (1992) definition, but the effect is modest.[16]

Taken together, our Bayesian meta-regression analysis uncovers some factors that are associated with the size of reported loss aversion coefficients, but it is still a difficult task to draw a complete picture of the observed heterogeneity. We note that 15.5% of the between-observation variance is explained by covariates.

4.5 Publication Bias

The cumulation of scientific knowledge is threatened by selective reporting or publication of findings. For example, suppose a theory or body of evidence makes a strong prediction about the sign or magnitude of a certain effect. Selective reporting occurs if scientists, editors and reviewers believe effects to be the norm, and there is a bias against reporting or publishing "unusual" results which contradict the norm. Selective reporting of this kind slows down the crucial process of scientific self-correction.

We will refer to such selective reporting of scientific findings collectively as "publication

[16]We have 10 estimates of λ reported in eight papers that use the definition of loss aversion according to Kőszegi and Rabin (2006, KR). Since the KR formulation incorporates consumption utility (as we discuss in Section 2), caution is needed when comparing estimates of λ using the KR formulation and other estimates of λ following the standard Tversky and Kahneman's (1992) definition. DellaVigna (2018), for example, suggests that a loss aversion of 2.25 in Tversky and Kahneman (1992) translates into a loss aversion of 3.35 in KR (footnote 16, p. 674).

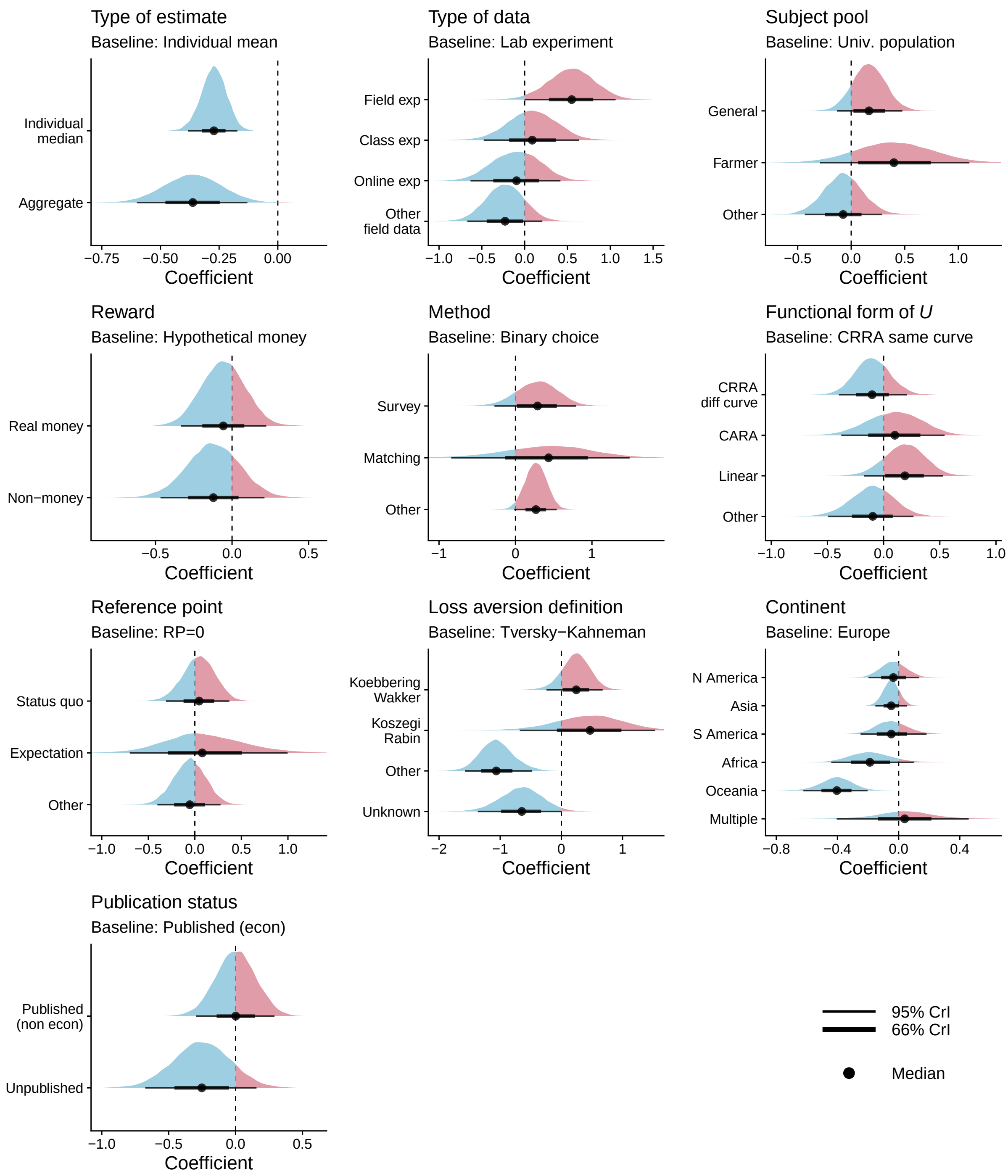

FIGURE 8: Bayesian random-effects meta-regression. Posterior distributions of coefficients β, together with posterior medians (black dot), 66% (thick solid line) and 95% (thin solid line) credible intervals, are shown. Table D.1 in the Online Appendix presents the result in a table format.

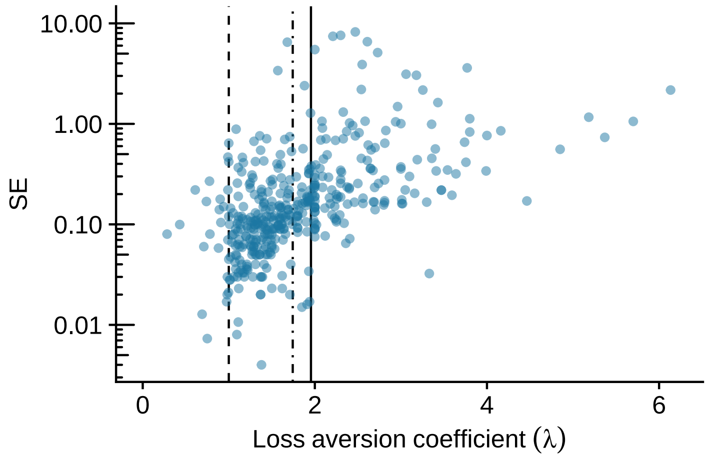

FIGURE 9: Funnel plot. *Notes*: The vertical dashed line corresponds to no loss aversion $\lambda = 1$. The vertical solid line corresponds to the estimated mean λ_0 from model M2, 1.955. The vertical dash-dotted line corresponds to the bias-corrected mean λ from model M1-ext, 1.742. The x-axis is cut off at 6 and the y-axis is displayed in the logarithmic scale for better visualization.

bias." Publication bias can take on many forms. We discuss two possibilities of relevance for the case of loss aversion. The first case is when journals prefer to see numerical estimates of loss aversion close to a certain number, and are skeptical of estimates that deviate far from that number. For there to be a bias, the preferred estimate by journal editors must differ from the "true" estimate found in the population of studies (see Borenstein et al., 2009).

A second form of bias concerns the overall significance of results. In this form, journals prefer results that are "statistically significant" (usually demarcated by a p-value of 0.05) and reject some null hypothesis (Andrews and Kasy, 2019; Brodeur et al., 2016; Brodeur, Cook and Heyes, 2020; Chopra et al., 2022). From the viewpoint of the authors of this paper, there is not a single specific test that can address these forms of publication bias jointly. Even if there were, the particular issues with this data might make it difficult to trust a single test absolutely. Instead we will take a descriptive approach to both types of publication bias.

Regarding the first form, in the context of loss aversion, one might suspect that researchers preferentially report evidence for loss aversion ($\lambda > 1$) and put evidence for loss tolerance ($\lambda < 1$) "in the file drawer" because such results contradict the initial hypothesis (Rosenthal, 1979). Other sources of publication bias are possible. Researchers in some disciplines may be motivated to undermine the "prevailing paradigm" of loss aversion and preferentially create or publish low-λ results. Alternatively, some journals may be interested in publishing results that conform to the pre-prospect theory economic orthodoxy of no loss aversion. Even coming up with a null hypothesis is thus more complex in our case than it would be when trying to simply ascertain the effect of a treatment or its absence.

Funnel plots (see Section 4.2) are often used as a tool to examine bias from the results of meta-analyses (Egger et al., 1997; Stanley and Doucouliagos, 2010). Figure 9 shows such a plot for all data points for which we have both an estimate of loss aversion and an associated standard error (thus excluding estimates with imputed SEs). In the absence of publication bias, the observations at the bottom of the graph, which have higher precision, should be be concentrated around the underlying mean estimate, indicated by the solid vertical line. As we move up in the graph toward the top of the funnel, and the precision of the studies decreases, we expect an increase in the degree of dispersion around the mean estimate. In the absence of publication bias, this dispersion ought to be symmetric around the mean. A larger number of observations in the upper right side of the graph compared to the upper left side would then be an indicator of classical publication bias, whereby estimates of loss aversion that fall closer to 1 and are not significant are less likely to be reported.

At first sight, there would indeed appear to be such an asymmetry in the graph. We clearly observe some large estimates in the upper right part of the graph, and hardly any corresponding estimates in the upper left part. Even more pronounced is the large cluster of studies at the bottom left corner of the graph. One might be concerned they indicate a model of publication bias where the true λ is around 0.8–1 but editorial bias favors publication of studies with higher estimates at lower precision. (In such a model, low-precision, low-estimate studies would not be published because of the judgment of journal referees or editors, and high-precision, high-estimate studies would not be published because they are statistically improbable.)

To test asymmetry in the funnel plot through examining correlation between reported estimates and their SEs, we extend the benchmark model M1 by allowing the underlying mean for each reported estimate to depend systematically on the standard error:

$$\lambda_i \mid \lambda_0, \gamma, \tau, se_i \sim \mathcal{N}\left(\lambda_0 + \gamma\sqrt{\tau^2 + se_i^2}, \tau^2 + se_i^2\right), \tag{M1-ext}$$

where γ represents a potential publication bias and $\lambda_0 + \gamma\tau$ captures the average loss aversion coefficient correcting for the bias (since $\mathbf{E}[\lambda_i] \to \lambda_0 + \gamma\tau$ as precision goes to infinity, or equivalently, $se_i \to 0$).[17] We find that the estimated mean γ is positive (i.e., 1.434 with a 95% CrI of $[1.008, 1.876]$), consistent with publication bias of a specific form: higher λ are less-precisely estimated and "small" λ are hidden in the literature. The "corrected" mean λ is the mean $\lambda_0 + \gamma\tau$ which is 1.742 with a 95% CrI of $[1.674, 1.811]$. Thus, the model suggests a relatively milder publication bias than Figure 9 might suggest, but nonetheless a drop from the average λ of our main estimate of 1.955 (model M2, Table 5).

[17] This model is motivated by the regression-based approach for detecting and correcting for publication bias in meta-analysis, introduced first by Egger et al. (1997) and established by Card and Krueger (1995), Stanley (2005, 2008), and Stanley and Doucouliagos (2014). Here we follow the "extended random-effects model" developed by Rücker et al. (2011), to be consistent with the hierarchical model we set up in Section 4.3.

The use and appropriate interpretation of funnel plot asymmetry has been debated, and there are reasons to not take this estimate at face value. A crucial assumption in measuring publication bias using these tests is that there is no correlation between estimates and their standard errors (in the absence of publication bias and between-study heterogeneity). In the context of estimation of behavioral parameters in experimental economics like ours, there are plausible reasons why the no-correlation assumption might fail. First, it is possible that researchers choose parameters in their experiments (such as a series of monetary outcomes used in a Multiple Price List) in a way that is tuned to detect loss aversion coefficient that are close to 1 or 2. Second, the parameter λ must be larger than 0 by construction of the theory, and the reported estimates exhibit non-normality. In each of these cases, the correlation between estimates and standard errors can arise "mechanically" even without any publication bias.[18]

A different way to examine a certain type of publication bias is to compare values reported in published and unpublished papers. This method does not rely on assumed independence between estimates and their standard errors. Publication bias by journals implies that the estimates found in published papers could vary from estimates in unpublished working papers. Our meta-regressions (presented in Section 4.4) estimated that observed loss aversion coefficients were 0.25 lower in working papers than in published studies, roughly 1.73 for working papers (though the two credible intervals are overlapping). Remember that the "true" value of lambda is best estimated by a weighted average of the working paper estimate and the published paper estimate, so 1.73 can still be thought of as a lower bound of the "true" estimate. It is perhaps reassuring that his value is almost the same as the bias-corrected estimate of 1.74 discussed above.

A second form of publication bias involves a journal focusing too much on "statistically significant" results. In this form, journals prefer to publish results that reject the null hypothesis for the parameter of loss aversion, such as $\lambda = 1$ or $\lambda = 2$, which could result in a distribution of test statistics that exhibits a discontinuity around a threshold for statistical significance (such as $z = 1.96$). Gerber and Malhotra (2008*a*,*b*) introduced a "caliper test" to identify a systematic bunching in the distribution of test statistics within narrow bands around threshold of statistical significance, but such an approach requires more observations than we have in our sample. Absent the ability to perform that test, we provide histograms of the z-statistics of our estimates in Figure 10. There are two histograms, centered around the null hypotheses of $\lambda = 1$ (panel A) and $\lambda = 2$ (panel B). The figure does not appear to reveal any such spikes in the positive or negative direction around values of 1 or 2 (which would appear around the vertical dotted line representing zero deviation), regardless of which null

[18] A similar point was raised by Matoušek, Havránek and Iršová (2022), who did a meta-analysis of experimentally measured individual discount rates which are typically bounded at zero. See also Sterne et al. (2011) for more general discussion and recommendations for interpretation of funnel plot asymmetry in the context of meta-analyses in RCTs.

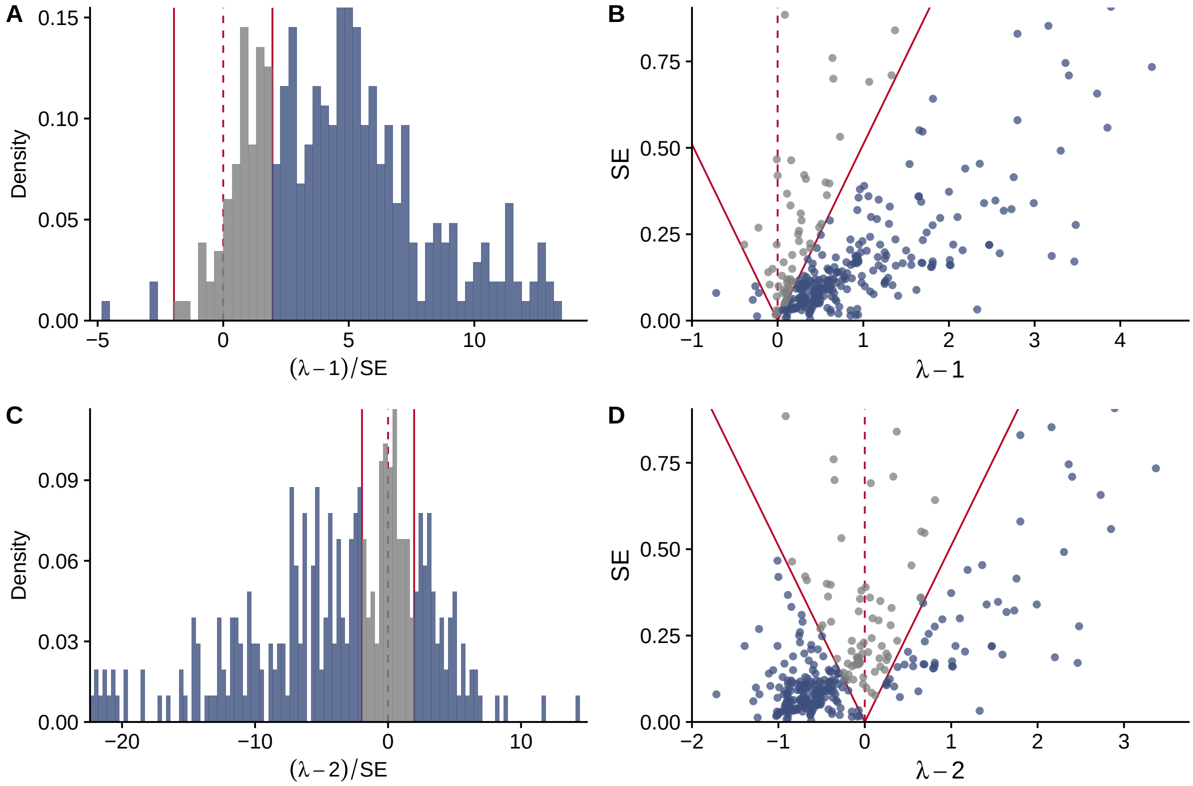

FIGURE 10: The left panels (AC) show binned density plots for the z-statistics and the right panels (BD) plot the estimated "effect size" against its standard error. In panels AB, the effect size is $\lambda - 1$, corresponding to the null hypothesis of $H_0 : \lambda = 1$. In panels CD, the effect size is $\lambda - 2$, corresponding to the null hypothesis of $H_0 : \lambda = 2$. *Notes*: The plots include 346 observations of aggregate-level and individual-level mean λ, which have associated standard errors reported in the paper. The solid red lines mark $|z| = 1.96$. Blue bars and dots correspond to effect sizes that are significantly different from zero at the 5% level. The bin width is 0.32 in panels AC. Outliers are not shown in the plots: the x-axis is restricted to the interval $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$ in panels AC and both axes are cut off at $Q3 + 3 \times IQR$ in panels BD. Figure D.8 in the Online Appendix shows the "full" version of panels BD, including outliers.

hypothesis we consider.

5 Discussion

Loss aversion is an important concept in behavioral economics and has been applied widely. This paper reports a meta-analysis of empirical estimates of the loss aversion coefficient λ. Our preferred specification indicates a mean $\lambda = 1.955$ and a 95% credible interval of $[1.820, 2.105]$. Many other specifications are within 0.1–0.3 of this finding and produce credible intervals that do not include 1 or 2.25. The former number is consistent with no loss aversion; the latter is an early estimate from Tversky and Kahneman (1992) which seems a bit too high. While there is a wide degree of heterogeneity across estimates, in general, no single factor emerges from meta-regression that greatly changes estimated loss aversion. Estimates derived from non-university populations, field experiments, and means of individual elicitations (compared to aggregates) are correlated with a modest increase in the loss aversion parameter.

A main takeaway from this paper is that the point estimate of $\lambda = 1.955$ reported above—estimated jointly with the uncertainty surrounding it—is the best current answer to the question of how large the loss aversion coefficient truly is. It is not the ultimate answer, however. One can indeed easily update this information with new evidence, either by using this as a Bayesian (hyper-)prior in subsequent estimations of loss aversion, or by combining the posterior from our meta-analysis with equivalent evidence from a follow-up meta-analysis on studies not yet included into our sample, e.g., because they appeared after our cut-off date.

We note two possible limitations to our study and, wherever possible, note where successive studies could improve. First, a key concern with any empirical analysis is differential selection of reported data. The outlined criteria for inclusion in our dataset are explicit and objective; we have no reason to believe, ex-ante, it should be correlated with our parameter of interest. However, our dataset is dominated by published studies. To the extent that publication may vary with a reported λ parameter, our analysis may suffer from bias.[19] Because the parameter of interest here is bounded by zero and positively skewed, it is difficult to use standard meta-analysis techniques like funnel plot (see Figure 9) and regressions to measure publication bias. Further, an important test of publication bias is whether unpublished working papers report reliably different results than published ones. There appears to be a modest difference; unpublished papers are about 0.25 lower which, notably, is still within the credible interval of our current estimate, and may be due to chance given the small sample. While our preferred estimate is $\lambda = 1.955$, based on the assumption that available studies are unbiased, researchers more concerned about publication bias may want to consider lower values (e.g., λ = 1.73-1.74).

Second, while we followed conventional practices of coding study quality (see Section 3.2), medical and health meta-analyses follow a more formal method to assess study quality as "risk of bias." Tens of thousands of Cochrane meta-analyses and systematic reviews follow a standard protocol where multiple raters evaluate risks of bias (study quality) on multiple dimensions such as the randomness of the treatment assignment and the blinding of participants and researchers, among others.[20] For instance, Hollands et al. (2015) is a meta-analysis

[19]It is important to note that any research synthesis technique would suffer from this issue, not just meta-analysis (Borenstein et al., 2009).

[20]Risk of Bias, introduced in 2008, is a tool widely used in Cochrane reviews and other systematic reviews for assessing risk of bias in randomized trials (Higgins et al., 2011). Their general criteria are: Randomness, blinding of participants and personnel (to treatment), blinding of outcome assessment, incomplete outcomes (attrition), selective reporting, comparability of groups (a.k.a. balance check), and consistency of intervention delivery. In each of these categories there are a small number of precise questions which are subjectively rated by assessors

of experiments testing the causal effect of changing portion, package, or tableware size on how much food, alcohol, or tobacco is chosen or actually consumed. Their Figure 3 shows that of their 72 studies assessed about one-third were Low Risk, the modal percentage were Unclear Risk, and less than 10% were deemed High Risk.[21]

Meta-regression analysis is uniquely able to adjust for unobserved differences in research quality across studies. If every study in our dataset contained multiple estimates, a fixed-effect panel meta-regression would have eliminated any potential distortion that differences is study quality may have on research findings (Stanley and Doucouliagos, 2012, pp. 112-117). Outside of this special case, the meta-regression can only control for effects and is dependent on how the values of quality are coded. Accordingly, we have recorded for each estimate the impact factor of the journal in which they were published (if applicable). The impact factor value has little explanatory power in regards to our results. We note (1) there is no meaningful correlation between journal impact factor and reported (mean) λ value; (2) inclusion or exclusion of journal impact factor does not meaningfully change our meta-regression results. To the extent journal impact factor can be a proxy for study quality, we see no evidence to support the concern that "low" quality studies are biasing our results.

In the introduction we promised to return to subtle, important comparisons between meta-analysis and narrative review. We will do that now.

Many readers will not know much about meta-analysis and might be skeptical about it, especially in comparison to the familiar style we call "narrative review." We anticipate such skepticism and offer counter-arguments. The unabashed goal is to advocate for more appreciation of the underdog method of meta-analysis.

We will start with fears about narrative review. Such reviews could be influenced by biases in remembering recent (and perhaps socially-connected) salient, recently-encountered data. Meta-analysis is a partial antidote.

There are two natural fears about meta-analysis. One is that it is somehow a mistake to reduce all research in a field to a single value or a range of numbers. The second fear is that simple apples-to-apples comparisons across studies neglects heterogeneity of methods (see Borenstein et al., 2009, for a history of such critiques). These fears are natural, but meta-analysis techniques have evolved to allay such fears.

The "single number" fear is misguided because humans—including scientists—value sim-

on a four-point categorical scale (Yes, Probably Yes, Probably No, No) then aggregated. See Sterne et al. (2019) for a revised tool (RoB 2).

[21] A possible analogue exists within social science research (especially for experiments), forecasts of the likely replicability of a study's findings. Altmejd et al. (2019) use data from actual replications to fit a machine learning model. They are able to predict the degree of replicability with about 75% accuracy from simple observable features of a study.

plicity. If meta-analysis did not provide a carefully and transparently derived value (or range), researchers will imagine another simplified value, one way or another. It is also often necessary to choose *some* value as an input into structural models within economics, and to make power calculations during pre-registration. Furthermore, the purpose of this or any meta-analysis is not to simply provide one number, but to demonstrate how such numbers vary given other factors that categorize studies, using meta-regression techniques (e.g., Figure 8). Stanley and Doucouliagos (2012) remind us that meta-regressions simply take a well-known and established technique to understand variation in many kinds of economic data, and just apply it to the data our own profession generates.

Next we turn to a sharper comparison between meta-analysis and narrative review. In a narrative review, there is no explicit attempt to canvas all studies based on stated criteria. Instead, an expert reviewer chooses studies that seem to be of especially high quality or pivot the scientific trajectory in a useful new direction. It is similar to an historical analysis of progress in a field. To relate the two methods, a narrative review is simply a meta-analysis with an altered subjective weighting system. The subjective weights are judgments by the reviewer of what findings readers should care the most about, and the reasons why.

One analogy is to sports commentary. Meta-analysis is like a "play-by-play" analyst who describes every action on the field with an even tone. Narrative review is like a "color analyst" who picks out certain plays which are unusually important and explains why they are special. The color analyst *adds* the dramatic emphasis that the play-by-play analyst suppresses. The analogy should make clear why both kinds of commentary are useful; the two together are better than each one alone.[22] A similar analogy is to the proud divide in newspapers between the "news" side (meta-analysis) and "opinion" (narrative review).

A tricky, interesting question is what meta-analysis and narrative review can say about influential "breakthrough" papers? Narrative reviews often remark on how a particular study represents a breakthrough in using new methods or presents a surprising finding that should be prioritized to be studied further. Because meta-analysis is backward-looking, it is not ideally-equipped to identify useful breakthroughs. Because narrative review is subjective, it can look forward and might do better.

An example is De Martino, Camerer and Adolphs (2010). They found that two patients with damage to the amygdala area of the brain were not averse to losses at all. This is a tiny finding with a large standard error; it's a teaspoon of water added to a swimming pool. Meta-regression is worthless because there is no power to detect an "amygdala damage" vs. "no amygdala damage" difference if the study were added to our corpus. But despite the tiny

[22] Another useful analogy comes from performance evaluation in personnel economics. It is well-known Baker, Gibbons and Murphy (1994) that objective and subjective measure of performance can be complements. Objective measures rein in over-the-top subjective evaluation, and subjective evaluations can add in idiosyncratic information that objective measures are blind to.

$n = 2$, it is a clue that could shed light on the fundamental mechanism underlying loss aversion, and hence deserves further study. Narrative reviews may miss some opportunities to amplify such interesting studies too, but meta-analysis will always neglect them.

We will conclude with one idea about how meta-analysis can guide future research. Meta-regression can actually pinpoint where studies are plentiful and where an additional study would have the greatest new effect on collective knowledge. A low standard error on a meta-regression coefficient means we do not need to learn more. A high standard error means that we do need to learn more. From our dataset, studies (i) other than lab and field experiments, (ii) focusing on specific, non-University student populations, (iii) on continents of South America, Africa and Oceania, (iv) involving rewards not expressed in monetary terms, (v) obtaining preferences in methods other than sequential binary methods, (vi) using utility functions other than CRRA with equal curvature for gains and losses, and (vii) with loss aversion specifications other than Kahneman and Tversky are the areas where we have uncovered the least data. Interesting new findings about the loss aversion parameter are *ex ante* more likely to be in those areas. The old cliché "we encourage future research in these areas" now has empirical backing.

References

Abdellaoui, Mohammed, Han Bleichrodt, and Corina Paraschiv. 2007. "Loss Aversion Under Prospect Theory: A Parameter-Free Measurement." *Management Science*, 53(10): 1659–1674.

Abdellaoui, Mohammed, Han Bleichrodt, and Olivier l'Haridon. 2008. "A Tractable Method to Measure Utility and Loss Aversion under Prospect Theory." *Journal of Risk and Uncertainty*, 36(3): 245–266.

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond. 2019. "Preferences for Truth-Telling." *Econometrica*, 87(4): 1115–1153.

Alesina, Alberto, and Francesco Passarelli. 2019. "Loss Aversion in Politics." *American Journal of Political Science*, 63(4): 936–947.

Altmejd, Adam, Anna Dreber, Eskil Forsell, Juergen Huber, Taisuke Imai, Magnus Johannesson, Michael Kirchler, Gideon Nave, and Colin Camerer. 2019. "Predicting the Replicability of Social Science Lab Experiments." *PLOS ONE*, 14(12): e0225826.

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and Correction for Publication Bias." *American Economic Review*, 109(8): 2766–2794.

Attema, Arthur E., Werner B. F. Brouwer, and Olivier l'Haridon. 2013. "Prospect Theory in the Health Domain: A Quantitative Assessment." *Journal of Health Economics*, 32(6): 1057–1065.

Baker, George, Robert Gibbons, and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, 109(4): 1125–1156.

Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang. 2016. "Prospect Theory and Stock Returns: An Empirical Test." *Review of Financial Studies*, 29(11): 3068–3107.

Barberis, Nicholas, and Ming Huang. 2001. "Mental Accounting, Loss Aversion, and Individual Stock Returns." *Journal of Finance*, 56(4): 1247–1292.

Barberis, Nicholas, and Ming Huang. 2008. "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices." *American Economic Review*, 98(5): 2066–2100.

Barberis, Nicholas, and Wei Xiong. 2009. "What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation." *Journal of Finance*, 64(2): 751–784.

Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives*, 27(1): 173–96.

Barberis, Nicholas, Lawrence Jin, and Baolian Wang. 2021. "Prospect Theory and Stock Market Anomalies." *Journal of Finance*, 76(5): 2639–2687.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics*, 116(1): 1–53.

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, 110(1): 73–92.

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis.* West Sussex, United Kingdom:John Wiley & Sons.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics*, 8(1): 1–32.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review*, 110(11): 3634–3660.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *Quarterly Journal of Economics*, 112(2): 407–441.

Card, David, and Alan B. Krueger. 1995. "Time-Series Minimum-Wage Studies: A Meta-Analysis." *American Economic Review*, 85(2): 238–243.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell.

2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, 76(1).

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

Chateauneuf, Alain, and Peter Wakker. 1999. "An Axiomatization of Cumulative Prospect Theory for Decision Under Risk." *Journal of Risk and Uncertainty*, 18(2): 137–145.

Chopra, Felix, Ingar Haaland, Christopher Roth, and Andreas Stegmann. 2022. "The Null Result Penalty." CEPR Discussion Paper No. 17331.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47(2): 315–372.

DellaVigna, Stefano. 2018. "Structural Behavioral Economics." In *Handbook of Behavioral Economics - Foundations and Applications*. Vol. 1, ed. B. Douglas Bernheim, Stefano DellaVigna and David Laibson, 613–725. Amsterdam:Elsevier/North-Holland.

De Martino, Benedetto, Colin F. Camerer, and Ralph Adolphs. 2010. "Amygdala Damage Eliminates Monetary Loss Aversion." *Proceedings of the National Academy of Sciences*, 107(8): 3788–3792.

Doucouliagos, Chris, Tom D. Stanley, and Margaret Giles. 2012. "Are Estimates of the Value of a Statistical Life Exaggerated?" *Journal of Health Economics*, 31(1): 197–206.

Doucouliagos, Hristos, and Tom D. Stanley. 2009. "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations*, 47(2): 406–428.

Doucouliagos, Hristos, Tom D. Stanley, and W. Kip Viscusi. 2014. "Publication Selection and the Income Elasticity of the Value of a Statistical Life." *Journal of Health Economics*, 33: 67–75.

Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal*, 315(7109): 629–634.

Engel, Christoph. 2011. "Dictator Games: A Meta Study." *Experimental Economics*, 14(4): 583–610.

Ericson, Keith M. Marzilli, and Andreas Fuster. 2014. "The Endowment Effect." *Annual Review of Economics*, 6: 555–579.

Fishburn, Peter C., and Gary A. Kochenberger. 1979. "Two-Piece von Neumann-Morgenstern Utility Functions." *Decision Sciences*, 10(4): 503–518.

Fryer, Roland G., Steven D. Levitt, John List, and Sally Sadoff. 2012. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment." NBER Working

Paper No. 18237.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2014. *Bayesian Data Analysis.* Boca Raton:CRC Press.

Gerber, Alan, and Neil Malhotra. 2008*a*. "Do Statistical Reporting Standards Affect What is Published? Publication Bias in Two Leading Political Science Journals." *Quarterly Journal of Political Science*, 3(3): 313–326.

Gerber, Alan S., and Neil Malhotra. 2008*b*. "Publication Bias in Empirical Sociological Research: Do Arbitrary Significance Levels Distort Published Results?" *Sociological Methods & Research*, 37(1): 3–30.

Glass, Gene V. 1976. "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher*, 5(10): 3–8.

Gurevitch, Jessica, Julia Koricheva, Shinichi Nakagawa, and Gavin Stewart. 2018. "Meta-Analysis and the Science of Research Synthesis." *Nature*, 555(7695): 175–182.

Havránek, Tomáš. 2015. "Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting." *Journal of the European Economic Association*, 13(6): 1180–1204.

Havránek, Tomáš, Marek Rusnak, and Anna Sokolova. 2017. "Habit Formation in Consumption: A Meta-Analysis." *European Economic Review*, 95: 142–167.

Havránek, Tomáš, Roman Horváth, Zuzana Iršová, and Marek Rusnak. 2015. "Cross-Country Heterogeneity in Intertemporal Substitution." *Journal of International Economics*, 96(1): 100–118.

Heidhues, Paul, and Botond Kőszegi. 2018. "Behavioral Industrial Organization." In *Handbook of Behavioral Economics - Foundations and Applications*. Vol. 1, ed. B. Douglas Bernheim, Stefano DellaVigna and David Laibson, 517–612. Amsterdam:Elsevier/North-Holland.

Higgins, Julian P. T., Douglas G. Altman, Peter C. Gøtzsche, Peter Jüni, David Moher, Andrew D. Oxman, Jelena Savović, Kenneth F. Schulz, Laura Weeks, and Jonathan A. C. Sterne. 2011. "The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials." *British Medical Journal*, 343: d5928.

Hollands, Gareth J., Ian Shemilt, Theresa M. Marteau, Susan A. Jebb, Hannah B. Lewis, Yinghui Wei, Julian P. T. Higgins, and David Ogilvie. 2015. "Portion, Package or Tableware Size for Changing Selection and Consumption of Food, Alcohol and Tobacco." *Cochrane Database of Systematic Reviews*, 9: CD011045.

Imai, Taisuke, Tom A. Rutter, and Colin F. Camerer. 2021. "Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets." *Economic Journal*, 131(636): 1788–1814.

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

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* New York:Farrar, Straus and Giroux.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, 47(2): 263–292.

Kim, E. Han, Adair Morse, and Luigi Zingales. 2006. "What Has Mattered to Economics Since 1970." *Journal of Economic Perspectives*, 20(4): 189–202.

Köbberling, Veronika, and Peter P. Wakker. 2005. "An Index of Loss Aversion." *Journal of Economic Theory*, 122(1): 119–131.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4): 1133–1165.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review*, 97(4): 1047–1073.

Kruschke, John K. 2010. "Bayesian Data Analysis." *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(5): 658–676.

l'Haridon, Olivier, and Ferdinand M. Vieider. 2019. "All Over the Map: A Worldwide Comparison of Risk Preferences." *Quantitative Economics*, 10(1): 185–215.

Matoušek, Jindšich, Tomáš Havránek, and Zuzana Iršová. 2022. "Individual Discount Rates: A Meta-Analysis of Experimental Evidence." *Experimental Economics*, 25(1): 318–358.

Meager, Rachael. 2022. "Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature." *American Economic Review*, 112(6): 1818–1847.

Mrkva, Kellen, Eric J. Johnson, Simon Gächter, and Andreas Herrmann. 2020. "Moderating Loss Aversion: Loss Aversion Has Moderators, but Reports of Its Death are Greatly Exaggerated." *Journal of Consumer Psychology*, 30(3): 407–428.

Neumann, Nico, and Ulf Böckenholt. 2014. "A Meta-Analysis of Loss Aversion in Product Choice." *Journal of Retailing*, 90(2): 182–197.

O'Donoghue, Ted, and Charles Sprenger. 2018. "Reference-Dependent Preferences." In *Handbook of Behavioral Economics - Foundations and Applications.* Vol. 1, ed. B. Douglas Bernheim, Stefano DellaVigna and David Laibson, 1–77. Amsterdam:Elsevier/North-Holland.

Oosterbeek, Hessel, Randolph Sloof, and Gijs van de Kuilen. 2004. "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis." *Experimental Economics*, 7(2): 171–188.

Pearson, Karl. 1904. "Report on Certain Enteric Fever Inoculation Statistics." *British Medical Journal*, 2(2288): 1243–1246.

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–1292.

R Core Team. 2020. "R: A Language and Environment for Statistical Computing." Vienna, Austria, R Foundation for Statistical Computing. https://www.R-project.org/.

Rieger, Marc Oliver, Mei Wang, and Thorsten Hens. 2017. "Estimating Cumulative Prospect Theory Parameters from an International Survey." *Theory and Decision*, 82(4): 567–596.

Romer, Paul M. 1996. "Preferences, Promises, and the Politics of Entitlement." In *Individual and Social Responsibility: Child Care, Education, Medical Care, and Long-Term Care in America*. 195–228. University of Chicago Press.

Rosenthal, Robert. 1979. "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin*, 86(3): 638–641.

Rücker, Gerta, Guido Schwarzer, James R. Carpenter, Harald Binder, and Martin Schumacher. 2011. "Treatment-Effect Estimates Adjusted for Small-Study Effects via a Limit Meta-Analysis." *Biostatistics*, 12(1): 122–142.

Schmidt, Ulrich, and Horst Zank. 2005. "What Is Loss Aversion?" *Journal of Risk and Uncertainty*, 30(2): 157–167.

Stan Development Team. 2020. "RStan: the R interface to Stan." R package version 2.21.2. http://mc-stan.org/.

Stanley, Tom D. 2001. "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives*, 15(3): 131–150.

Stanley, Tom D. 2005. "Beyond Publication Bias." *Journal of Economic Surveys*, 19(3): 309–345.

Stanley, Tom D. 2008. "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics*, 70(1): 103–127.

Stanley, Tom D., and Hristos Doucouliagos. 2010. "Picture This: A Simple Graph That Reveals Much Ado About Research." *Journal of Economic Surveys*, 24(1): 170–191.

Stanley, Tom D., and Hristos Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business.* London:Routledge.

Stanley, Tom D., and Hristos Doucouliagos. 2014. "Meta-Regression Approximations to Reduce Publication Selection Bias." *Research Synthesis Methods*, 5(1): 60–78.

Stanley, Tom D., and Stephen B. Jarrell. 1998. "Gender Wage Discrimination Bias? A Meta-Regression Analysis." *Journal of Human Resources*, 947–973.

Sterne, Jonathan A. C., Alex J. Sutton, John P. A. Ioannidis, Norma Terrin, David R. Jones, Joseph Lau, James Carpenter, Gerta Rücker, Roger M. Harbord, Christo-

pher H. Schmid, et al. 2011. "Recommendations for Examining and Interpreting Funnel Plot Asymmetry in Meta-Analyses of Randomised Controlled Trials." *British Medical Journal*, 343: d4002.

Sterne, Jonathan A. C., Jelena Savović, Matthew J. Page, Roy G. Elbers, Natalie S. Blencowe, Isabelle Boutron, Christopher J. Cates, Hung-Yuan Cheng, Mark S. Corbett, Sandra M. Eldridge, et al. 2019. "RoB 2: A Revised Tool for Assessing Risk of Bias in Randomised Trials." *British Medical Journal*, 366: l4898.

Tovar, Patricia. 2009. "The Effects of Loss Aversion on Trade Policy: Theory and Evidence." *Journal of International Economics*, 78(1): 154–167.

Tunçel, Tuba, and James K. Hammitt. 2014. "A New Meta-Analysis on the WTP/WTA Disparity." *Journal of Environmental Economics and Management*, 68(1): 175–187.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.

Wakker, Peter, and Daniel Deneffe. 1996. "Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown." *Management Science*, 42(8): 1131–1150.

Wakker, Peter P. 2010. *Prospect Theory: For Risk and Ambiguity.* Cambridge:Cambridge University Press.

Walasek, Lukasz, Tim L. Mullett, and Neil Stewart. 2018. "A Meta-Analysis of Loss Aversion in Risky Contexts." Available at SSRN: http://dx.doi.org/10.2139/ssrn.3189088.

Wang, Mei, Marc Oliver Rieger, and Thorsten Hens. 2017. "The Impact of Culture on Loss Aversion." *Journal of Behavioral Decision Making*, 30(2): 270–281.

Yates, Frank, and William G. Cochran. 1938. "The Analysis of Groups of Experiments." *Journal of Agricultural Science*, 28(4): 556–580.