Meta-Analysis of Distributional Preferences*

Salvatore Nunnari*† Massimiliano Pozzi‡

November 28, 2025

Abstract

We conduct an interdisciplinary meta-analysis of empirical estimates of distributional preferences reported between 1999 and 2023, examining 297 estimates of sensitivity to inequality from 41 articles that structurally estimate the Fehr and Schmidt (1999) model. Our analysis indicates that individuals are inequality averse: mean sensitivity to disadvantageous inequality is 0.533; mean sensitivity to advantageous inequality is, instead, 0.326. We also uncover systematic heterogeneity: aversion to advantageous (disadvantageous) inequality is smaller (larger) in strategic environments, reflecting that these parameters bundle distributional motives with other motivational forces. Publication bias diagnostics reveal patterns compatible with selection, while p-curve evidence is inconsistent with p-hacking. Nonetheless, selection-corrected averages remain positive, economically large, and precisely estimated. An out-of-sample prediction exercise shows that our estimates outperform existing mixture models and can thus provide an efficient benchmark for applied work. Finally, we examine 98 estimates of altruism and attitude towards equity versus efficiency from 17 studies structurally estimating the Andreoni and Miller (2002) model. The representative individual displays Cobb–Douglas preferences with roughly 1/3 weight on others' earnings.

Keywords: Other-Regarding Preferences, Altruism, Envy, Guilt, Inequality Aversion, Meta-Analysis, Multi-Level Random-Effects Model, Bayesian Hierarchical Model
JEL Codes: C90, C11, D63, D91

---

*Corresponding Author: Salvatore Nunnari, salvatore.nunnari@unibocconi.it. Previous versions were circulated as "Meta-Analysis of Inequality Aversion Estimates" and "Meta-Analysis of Distributional Preferences Estimates." We are grateful to Jeffrey Carpenter, Lina Diaz, Daniel Houser, John Ifcher, Stephan Mueller, Sander Onderstal, Holger Rau, Andrea Robbett, Arthur Schram, Yang Yang, Silvia Vannutelli, and Homa Zarghamee for providing additional details on their estimates. We thank Sandro Ambuehl, David Cooper, Glenn Dutcher, Catherine Eckel, Ernst Fehr, Guillaume Frechette, Taisuke Imai, Daniel Mueller, Ferdinand Vieider, Roberto Weber, Rainer Winkelmann, and seminar audiences for helpful comments. We acknowledge financial support from the ERC (SN; Grant No. 852526) and the University of Zurich (MP; Research Program 'Equality of Opportunity').

†Bocconi University, Department of Economics; CEPR; CESifo; salvatore.nunnari@unibocconi.it.
‡University of Zurich, Department of Economics; massimiliano.pozzi@econ.uzh.ch.

1 Introduction

The standard economic model of choice assumes that individuals are only motivated by self-interest. In the last three decades, however, a large body of evidence from the experimental social sciences has showed that most people hold *other-regarding preferences*, that is, that they care about others' outcomes or whether others are treated fairly or not.

Other-regarding preferences have been successfully used to explain behavior which is commonly observed in laboratory experiments and field environments, yet puzzling from the perspective of the standard economic model of choice. With regard to laboratory findings, this includes responders' rejection of positive offers in ultimatum games (Güth, Schmittberger and Schwarze, 1982; Eckel and Grossman, 2001), proposers' positive offers in dictator games (Forsythe et al., 1994; Hoffman et al., 1994; Henrich et al., 2005), cooperation in the static prisoner's dilemma (Yamagishi and Kiyonari, 2000), positive contributions in the linear public good game (Ledyard, 1995), and positive amounts sent and returned in trust games (Berg, Dickhaut and McCabe, 1995; Burks, Carpenter and Verhoogen, 2003). With regard to behaviors in the field, this includes workers' negative effort responses to wage cuts (Bewley, 1999; Kube et al., 2013; Cohn et al., 2014; Kaur, 2019) and unequal payments (Cohn et al., 2014; Breza et al., 2018), workers' collective actions to suppress labor supply (Breza et al., 2019) and their quitting behavior in response to wage inequality (Dube et al., 2019), citizens participation in protest movements against dictatorships (Cantoni et al., 2022), investors investments in stocks that promise to care for the environment or for social responsibility (Riedl and Smeets, 2017) or the role of other-regarding preferences for redistributive politics (Kerschbamer and Müller, 2020; Fehr et al., 2024; Epper et al., 2024)

The most cited and influential model of other-regarding preferences is the model proposed by Fehr and Schmidt (1999) (FS henceforth).[1] In the simplest two-players version of this

[1] As of 2 June 2025, FS has 16,327 citations on Google Scholar and 6,838 citations on Web of Science.

model, the utility agent i derives from outcome x is

$$U_i(x) = x_i - \alpha_i \max[x_j - x_i, 0] - \beta_i \max[x_i - x_j, 0], \; j \neq i.$$

The agent's utility does not depend only on her own payoff, x_i, but also on the comparison with the other agent's payoff, x_j. Assuming that $\alpha > \beta > 0$ (as in FS), this is a model of *inequity aversion* (where α can be interpreted as *envy* and β as *guilt*), since differences in payoffs cause disutility for agent i. At the same time, this simple framework can capture other kinds of other-regarding preferences: if $\alpha < 0$ and $\beta < 0$, this is a model of *inequality seeking*; if $\alpha < 0$ and $\beta \geq 0$, this is a model of *altruistic preferences*; if $\alpha > 0$ and $\beta < 0$, this is a model of *spiteful preferences*; and if $\alpha < 0$ and $\beta > 0$, this is a model of *efficiency concerns.* This parsimonious utility specification is able to explain many of the above mentioned "anomalies" while keeping the model simple and tractable at the same time.

Despite all the work social scientists have done in the past 25 years to give the model an axiomatic foundation and to test it in the laboratory, there is still no consensus on what are plausible values of α and β in relevant populations. This information can be very valuable as it provides both empirical and theoretical researchers guidance on crucial modeling choices (i.e., preferences assumptions) that strongly affect predictions. Indeed, parameter estimates from small and peculiar samples have been used as benchmark in theoretical work with inequity averse agents to deliver counterfactuals and policy recommendations (see, e.g., Fehr and Schmidt 2004, Fehr, Klein and Schmidt 2007, Fehr, Kremhelmer and Schmidt 2008, Normann and Rau 2015, and Vogt 2016).[2] Moreover, models of distributional preferences have been used to explain or predict behavior with applications ranging from optimal

[2]In their original paper, FS calibrate a distribution of parameters to match the behavior observed in previous ultimatum game experiments (e.g., Roth and Erev 1995). This distribution assumes that α can take four different values in the population — 0, 0.5, 1 and 4 — with calibrated shares of, respectively, 30%, 30%, 30% and 10%; on the other hand, β was assumed to take three different values — 0, 0.25 and 0.6 — with calibrated shares of, respectively, 30%, 30% and 40%. More recently, Eckel and Gintis (2010) reviewed the mean parameters estimated in four studies other than FS and reported values ranging between 0.31 and 1.89 for α, and between -0.27 and 0.80 for β. Blanco, Engelmann and Normann (2011), instead, estimated the coefficients at the individual level using ultimatum and dictator games and reported average estimates of 1.18 for α and 0.47 for β.

climate policy (Azar and Sterner, 1996; Anthoff et al., 2009; Tol, 2010), optimal taxation (Aronsson and Johansson-Stenman, 2023), industrial organization (Huck et al., 2001) and trade protection (Lü et al., 2012), contract design (Fehr and Schmidt, 2004; Fehr et al., 2007, 2008), redistributive policies (Fehr et al., 2024), social choice (Le and Saporiti, 2024), and the backlash against globalization (Pástor and Veronesi, 2021). In all these cases, predictions hinge on what kind of distributional preferences economic or political agents are endowed with and it is important to use empirically validated assumptions.

In this paper, we aggregate the knowledge from empirical estimates of other-regarding preferences accumulated in over 25 years of research with the method of meta-analysis, that is, "the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings" (Glass, 1976). In a meta analysis, studies are selected using a precise inclusion criterion; then, the information contained in these studies is codified and summarized to explain both regularities and variation across studies.[3]

In particular, we collect 149 estimates of sensitivity to disadvantageous inequality and 148 estimates of sensitivity to advantageous inequality from 41 articles in economics, psychology, neuroscience and computer science which structurally estimate the FS model of social preferences. The unit of observation is either aggregate measures (that is, estimates for a "representative subject" in an experiment) or measures of central tendency (that is, means or medians for subjects participating to the same experiment) of the parameter estimates.[4] We use this novel dataset to tackle three research questions. First, given the accumulated knowledge, what is the best estimate of the average α and β? Second, how do the average α and β vary depending on the characteristics of a study (e.g., the experimental task and the

[3]Thus, meta-analysis differs from narrative reviews that give, instead, a descriptive overview of a research topic, presenting the historical trajectory and the key findings in the literature. While providing a useful summary of past research and suggesting future avenues, narrative reviews do not systematically analyze all studies asking the same research question in order to test a statistical hypothesis like meta-analyses do.

[4]This is the information commonly available in papers that structurally estimate the FS model and, thus, this allows us to collect the largest number of estimates. An important insight of the existing literature on distributional preferences (see, for example, Fehr and Charness 2025) is the role of individual heterogeneity. In Section 4.4, we discuss all the information on heterogeneity in individual-level estimates available in the studies we reviewed to create our dataset.

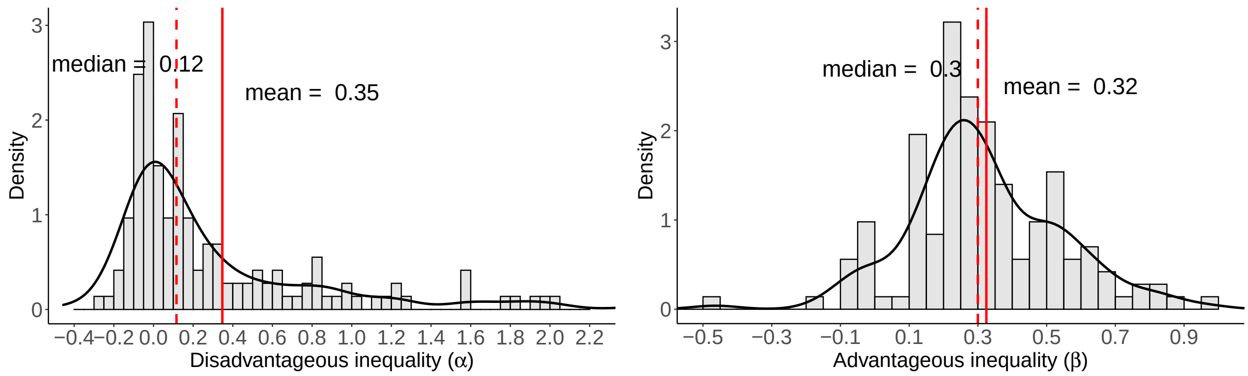


Figure 1: Distribution of Disadvantageous (α) and Advantageous Inequality (β) Coefficients. Notes: Bins for histograms are 0.05 wide; the Gaussian kernel density (solid black line) uses the Silverman's rule of thumb for bandwidth selection; in the panel for α, the horizontal axis is truncated at 2.2; in the panel for β, the horizontal axis is truncated at -0.5 and 0.9 for better visual rendering but the kernel density uses all estimates in both cases.

subject population)? Third, is there evidence of selective reporting or publication bias?

In order to answer the first question, we initially conduct a non-parametric analysis. Figure 1 shows the distribution of estimates in our dataset. The raw mean and median estimates of α are, respectively, 0.35 and 0.12 with around 30% of the estimates (45 out of 149) equal to or less than 0 (in contrast with the assumption in FS). The raw mean and median estimates of β are, respectively, 0.32 and 0.30, and, again, a sizeable number of observations which do not match the assumption in FS ($\beta \leq 0$ in 15 out of 148 estimates). Focusing on studies which estimate both parameters, disadvantageous inequality matters more than advantageous inequality only around 30% of the time (in 46 out of 144 pairs of estimates) and the correlation between the two parameters is indistinguishable from 0. In the non-parametric analysis, all estimates are given equal weight (even if the parameters computed in some studies are more reliable than others) and are assumed to be independent from one another (even if the same study provides multiple estimates). To tackle these issues, we compute a "weighted average" for α and β using a three-level random-effects model. Our analysis indicates that individuals are inequality averse: the meta-synthetic average for the disadvantageous inequality coefficient is 0.533, the meta-synthetic average for the advantageous inequality coefficient is 0.326, and both are strongly statistically significant.

The estimates in our dataset are drawn from studies that differ from one another, due to factors such as the subject population, the tasks participants engaged in, and other variables. Given this diversity, it seems unlikely that differences in α and β are simply driven by sampling errors. To explain this heterogeneity, we use the features of the studies and of the estimates we coded in our dataset as mediating variables. These meta-regressions reveal an interesting pattern: estimates of α computed using choices from strategic environments are larger than estimates computed using choices from individual decision-making tasks, while the reverse is true for estimates of β. In other words, strategic environments are associated with greater envy and smaller guilt.

This difference supports the view that the model may capture not only distributional concerns but also other psychological or motivational forces. First, greater envy in strategic environments suggests that efficiency motives (inducing individuals to internalize others' earnings) may be weakened by the competitive nature of strategic environments, where participants tend to view themselves as opponents rather than partners (as conjectured in Fehr, Naef and Schmidt 2006). Second, greater guilt in individual decision-making tasks could be due to a higher discomfort from a favorable comparison with others when the outcome is entirely attributable to one's own action and others only play a passive role. A deeper investigation of this correlations, where we unpack strategic environments into ultimatum games and other games, reveals that the greater envy documented in strategic environments is, at least in part, driven by responders' behavior in ultimatum games. This suggests that the estimate of α in strategic environments conflates concerns about equity with concerns about reciprocity or intentions (that are particularly salient for the responder in an ultimatum game). On the other hand, the smaller aversion to advantageous inequality in strategic environments is not driven by behavior in any particular game and, thus, confirms the role of the sense of responsibility towards others (which is lower in games) as conjectured by Fehr and Schmidt (2006) and Camerer (2003).

Finally, one aspect to keep in mind when conducting a meta-analysis is the problem of selective reporting and publication bias which arises when the probability of a study being published is affected by its results. In order to detect selective reporting and investigate the incidence of p-hacking, we use funnel plots, the Funnel Asymmetry Testing and Precision Effect Testing (FAT-PET) procedure (Stanley and Doucouliagos, 2012, 2017), histograms of z-statistics and the p-curve (Simonsohn, Nelson and Simmons, 2014). On the one hand, funnel plots highlight the absence of studies estimating (large in magnitude and imprecisely estimated) negative values of α and positive values of β and this is confirmed by the FAT-PET procedure. Moreover, we observe a jump around the threshold for statistical significance in the histograms of z-statistics for both parameters, which is a hint of p-hacking. On the other hand, the asymmetry in the funnel plots could be generated in the absence of publication bias — for example, because of feasibility constraints in the estimation of the parameters due to the experimental tasks employed or because of the implausible preferences implied by the missing values of α and β — and the publication-bias corrected meta-synthetic averages of the two parameters are still positive and strongly statistically significant (0.426 for α and 0.382 for β). In addition, the p-curves for both α and β are highly right-skewed which strongly supports the hypothesis that both parameters are different from zero and that researchers did not engage in p-hacking. We, thus, conclude that there is no compelling evidence of selective reporting or publication bias.

To the best of our knowledge, this is the first work that uses meta-analysis techniques to summarize empirical estimates of other-regarding preferences.[5] Our work builds on the narrative reviews on other-regarding preferences by Fehr and Schmidt (2006) and Cooper and Kagel (2016), the meta-analysis on dictator games by Engel (2011) and the meta-analysis on ultimatum games by Oosterbeek, Sloof and Van De Kuilen (2004) and Cooper and Dutcher (2011). These meta-analyses summarize the behavior observed in laboratory experiments testing ultimatum and dictator games and investigate the explanatory power of mediating

[5]Examples of meta-analyses in experimental and behavioral economics are Zelmer (2003) on linear public good games, Embrey, Fréchette and Yuksel (2018) on the finitely repeated prisoner's dilemma, Baranski and Morton (2021) on multilateral alternating-offer bargaining, Imai, Rutter and Camerer (2021) on time preferences, Brown et al. (2024) on loss aversion, and Meager (2019, 2022) on the effect of microcredit.

variables (e.g., the size of the pie and the location of the experiment) but do not discuss structural estimates of a model.

FS is not the only model of distributional preferences. In Section 5, we expand our systematic investigation of distributional preferences with a meta-analysis of parameter estimates from the Altruistic Constant Elasticity of Substitution (CES) Preferences model by Andreoni and Miller (2002) (AM henceforth). This is an influential framework that has been followed by many attempts to structurally estimate its parameters with experimental data and, thus, it is the ideal candidate for a meta-analysis.[6] Moreover, this analysis can provide complementary insights, as the AM model is meant to capture altruism and attitude towards the efficiency-equity tradeoff (rather than sensitivity to inequality as FS).

Andreoni and Miller (2002) assume the following utility function:

$$U(\pi_s, \pi_o) = [a\pi_s^{\rho} + (1-a)\pi_o^{\rho}]^{\frac{1}{\rho}}$$

where π_s is the allocation to self, π_o is the allocation to other, $a \in [0,1]$ is the relative weight on one's own payoff, and $\rho \in (-\infty, 1]$ is the curvature of indifference curves. Any $0 < \rho \leq 1$ indicates distributional preference weighted towards increasing total payoffs (i.e., efficiency) while any $\rho < 0$ indicates distributional preference weighted towards reducing payoff differences (i.e., equality). We examine 98 estimates from 17 articles in economics, psychology, and biology, and we show that the representative individual has Cobb-Douglas preferences over own and others' earnings (i.e., $\rho = 0$) with weight to others' earnings equal to 1/3 (i.e, $a = 2/3$).

The rest of this paper is organized as follows. Section 2 describes the FS model of distributional preferences. Section 3 describes how the data was assembled and coded. Section 4 presents the results. Section 5 discusses the meta-analysis of parameter estimates from the AM model. Section 6 concludes.

[6] As of 2 June 2025, AM has 2,714 citations on Google Scholar and 1,077 citations on Web of Science.

2 The FS Model of Other-Regarding Preferences

In this section, we describe the original model in FS, while we leave for Appendix K.1 a description of all the variations of the model whose parameters are structurally estimated by the studies in our dataset. Consider a set of N players indexed by i and a vector of outcomes (e.g., monetary payoffs), $x = (x_1, x_2, ..., x_N)$. FS assume that player i derives the following utility from x:

$$U_i(x) = x_i - \alpha_i \frac{1}{N-1} \sum_{j \neq i} \max[x_j - x_i, 0] - \beta_i \frac{1}{N-1} \sum_{j \neq i} \max[x_i - x_j, 0], \tag{1}$$

FS further assume $\alpha_i \geq \beta_i$ and $1 > \beta_i \geq 0$. With only two players, this simplifies to

$$U_i(x) = x_i - \alpha_i \max[x_j - x_i, 0] - \beta_i \max[x_i - x_j, 0], \ i \neq j. \tag{2}$$

The first term in equations (1) and (2) captures the utility from one's own outcome; the second term measures the disutility from being behind in pairwise comparisons (i.e., sensitivity to disadvantageous inequality); and the third term measures the disutility from being ahead in pairwise comparisons (i.e., sensitivity to advantageous inequality).

We briefly discuss the assumptions made in the original contribution by FS. First, FS assume $\alpha \geq 0$ and $\beta \geq 0$, making this a model of inequality aversion: fixing her own payoff, x_i, player i's utility is maximized when $x_j = x_i$ (see Figure 2). FS further assume that $\alpha \geq \beta$. This assumption implies that disadvantageous inequality hurts more than advantageous inequality and is inspired by earlier work in behavioral and experimental economics (Kahneman and Tversky, 1979; Loewenstein, Thompson and Bazerman, 1989). Finally, FS constrain β to be smaller than 1 in order to avoid an implausible scenario: agents with $\beta > 1$ are willing to burn money in order to reduce the favorable gap between their allocation and the allocation to others. As discussed in the Introduction, while this is interpreted as a model of inequality aversion when $\alpha > 0$ and $\beta > 0$, this parsimonious framework can be

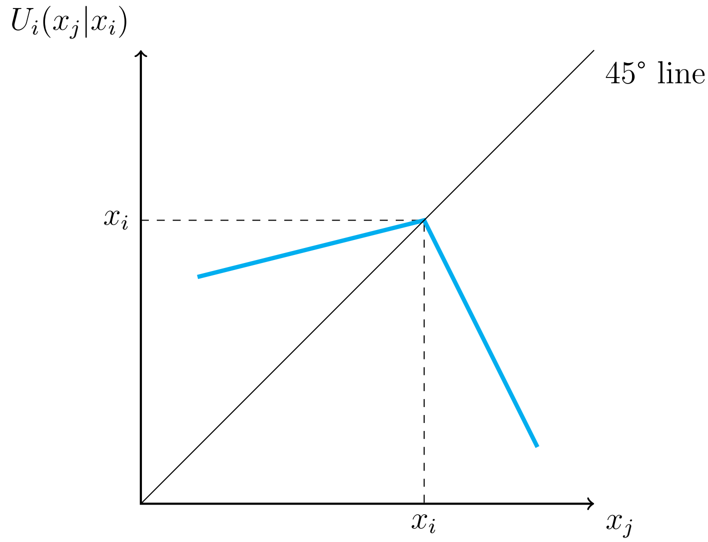


Figure 2: Utility of Inequality Averse Player i in Game with 2 Players ($\alpha = 2$, $\beta = 0.5$).

used to model different kinds of other-regarding preferences. Our meta-analysis will reveal which type of other-regarding preferences is more common in the populations that have been sampled in 25 years of social sciences experiments.

3 Data

3.1 Identification and Selection of Relevant Studies

In order to perform an unbiased meta-analysis, it is important to define a precise and unambiguous inclusion criterion. Our criterion is to include "all papers that estimated the parameters for sensitivity to disadvantageous inequality, α, and/or advantageous inequality, β, using the model by Fehr and Schmidt (1999)".[7]

The search procedure followed four steps. First, we read the narrative reviews by Fehr and Schmidt (2006) and Cooper and Kagel (2016) and searched on Google Scholar to find a first seed of papers that estimated α and β. Second, we read these papers to identify the best possible combination of keywords for a more detailed search. Third, we searched

[7]This definition includes also the models that use FS as baseline and augment it by adding other parameters as discussed in the Technical Appendix K.1.

''inequality aversion" OR ''inequity aversion" OR ''envy" OR ''guilt"
OR ''advantageous inequality" OR ''disadvantageous inequality"
OR ''advantageous inequity" OR ''disadvantageous inequity"
AND
(experiment* OR estimat* OR surve*)

Figure 3: Query Used for Search on Web of Science, Google Scholar and Scopus

the scientific citation indexing databases Web of Science (February 8, 2022), Google Scholar (February 8, 2022) and Scopus (7 September 2022) using the query in Figure 3. Since we are interested in estimates of the FS parameters, we restricted the search to papers that cite FS. This search returned 1,916 articles. We then read these articles and excluded papers that were clearly irrelevant for our analysis — for example, articles that measured other-regarding preferences in animals or studies that, while reporting the results of dictator and ultimatum games, did not estimate the parameters of interest. Finally, we read through the remaining articles and applied our inclusion criterion. The final dataset consists of 41 articles and the complete list is available in Appendix A.[8]

3.2 Data Construction

After identifying the relevant articles, we assembled the dataset for the meta-analysis by coding the estimates for α and β, the features of the studies and the features of the estimation methodology. The main variables of interest are the structural estimates for the two coefficients of sensitivity to advantageous and disadvantageous inequality. In our 41 articles, these estimates take four forms: (i) *aggregate*, where a single value for α and β is estimated for the pooled data of all subjects in the study; (ii) *finite-mixture*, where a finite number of values for α and β alongside their distributions are estimated from the pooled data of all subjects; (iii) *individual-level mean*, where α and β are estimated separately for each subject and the mean value of the parameters is reported; and (iv) *individual-level median*,

[8]When a precise measure of the estimated parameters was not available (e.g., because the article reported only a scatter plot or a bar chart of individual-level estimates), we contacted the authors to get additional details.

same as iii) but where the median (rather than the mean) is reported. The first, third and fourth types of estimates are ready to be used in the meta-analysis. For the finite-mixture estimates, we computed and coded a weighted average for each parameter.

The measure of estimation uncertainty is another important variable to code in the dataset. This information is fundamental when conducting a meta-analysis: instead of simply averaging estimates from various studies, our aggregation procedure gives more weight to estimates that have lower SEs and, thus, are more precisely estimated (for example, because they are computed from experiments with a larger sample size). Out of 297 estimates in our dataset, the source reported the SEs for 79 estimates and, in other 146 cases, we were able to compute the SEs using the reported standard deviation and sample size or from t-statistics. For the remaining 72 estimates, we did not have (direct or indirect) information about the SEs.[9]

We had two options: either drop the 72 estimates without SEs or approximate the SEs and keep these estimates in the dataset. We chose the latter option, especially since observations might not be dropped randomly, thus, introducing a bias in our results. For this reason, while using approximated SEs is a second-best, we deemed this as the more sensible option. Nonetheless, we present the main results of our meta-analysis both for the full sample and for the restricted sample that considers only estimates with reported (i.e., not approximated) SEs. For more details regarding the construction of the SEs see Appendix K.4.

Finally, we coded variables describing features of the studies and of the estimates. These variables include the paper publication status, the methodology (e.g., laboratory experiment, classroom experiment, online experiment), the subject population (e.g., non-representative sample of college students, non-representative sample of adults, sample representative of a target population), subjects' location of residence, the task used to elicit the parameters (e.g., dictator game, ultimatum game, etc.), the reward type, the utility function posited for

[9]This usually happens for articles that compute individual-level estimates but report only the mean or median without the standard deviation. In one case, the standard deviation was reported but the sample size was unclear.

the estimation (e.g., FS, FS plus Kantian morality, etc.), and the estimation method. The next subsection discusses the distribution of the main features in our dataset. The full list is available in Appendix B.

3.3 Features of Studies and Estimates in the Dataset

As discussed in Section 3.1, we identified 41 articles which estimated the advantageous and disadvantageous inequality parameters in FS. In our dataset, we use as unit of measure a single *study* rather than a single *paper*. These two objects usually coincide but there is one exception: Beranek, Cubitt and Gächter (2015) report results of three distinct laboratory experiments conducted in the UK, the US and Turkey with three different samples. In our terminology, each of these three laboratory experiments comes from the same paper but corresponds to a different study. This means that, overall, we have 43 studies (discussed in 41 papers). These studies report estimates for 153 models of social preferences à la FS: 144 models estimate both the advantageous and disadvantageous inequality parameters, 5 models estimate only α, and 4 models estimate only β.

Table 1 reports the coded features of the 43 studies in our dataset. Among the 43 studies, 40 were presented in papers published (as of June 2, 2025) in economics, psychology, neuroscience and computer science journals. The majority of these 43 studies conducted traditional in-person laboratory experiments, while 7 studies conducted experiments online.[10] The studies were conducted in 11 different countries (China, Denmark, France, Germany, Italy, Netherlands, Spain, Sweden, Switzerland, Turkey, UK, and US) and involved mostly college students (32 studies out of 43), with 4 studies recruiting a sample representative of the Danish, Dutch or German general population (Bellemare, Kröger and van Soest, 2008,

[10]One study recruited participants from mTurk, one from Prolific, two using CentERpanel (an online survey consisting of a representative sample of the adult Dutch population), one using the German Internet Panel (GIP, a longterm longitudinal study managed by the University of Mannheim and regularly interviewing online a representative sample of the adult German population), one using the internet Laboratory for Experimental Economics (iLEE) at the University of Copenhagen (with subjects selected to be a random sample from the general Danish population), and one contacting climate negotiators from the Intergovernmental Panel on Climate Change directly via email.

Table 1: Features of the Studies ($N = 43$) in the Dataset

	Frequency	Fraction
Publication Status		
Published	40	0.93
Unpublished	3	0.07
Methodology		
Laboratory Experiment	34	0.79
Classroom Experiment	1	0.02
Online Experiment	7	0.17
Multiple Methodologies	1	0.02
Geographic Location		
United States	11	0.26
Northern Europe (CH, DE, DK, NL, SE, UK)	20	0.46
Southern Europe (FR, IT, ES, TR)	6	0.14
China	3	0.07
Multiple or Unspecified Locations	3	0.07
Subject Population		
College Students	32	0.74
Non-Representative Sample of Adults	6	0.14
Representative Sample (of DE, DK, or NL)	4	0.10
Multiple Populations	1	0.02
Experimental Task Used To Estimate α		
Standard Dictator Game	3	0.07
Mini Dictator Game	3	0.07
Mini Dictator Game with Equality-Efficiency Trade-Off	19	0.44
Ultimatum Game	12	0.28
Other Game	11	0.26
Experimental Task Used To Estimate β		
Standard Dictator Game	3	0.07
Mini Dictator Game	3	0.07
Mini Dictator Game with Equality-Efficiency Trade-Off	27	0.63
Ultimatum Game	5	0.12
Other Game	11	0.26
Reward Type		
Money	43	1.00

Note: We label as 'Mini Dictator Game' a task where a single decision-maker chooses from a finite set of (exogenous) self/other allocations; in the papers, this task has different labels ('ultimatum game abstracted from strategic interactions', 'choice menu', 'equality equivalence test', 'inequality list', and 'random ultimatum game'). For a list of experimental tasks different than Ultimatum Games and Dictator Games see Appendix C. The frequencies for experimental tasks sum to a number greater than N because some studies use more than one task.

Table 2: Features of the Estimates ($N = 297$) in the Dataset.

	α ($N = 149$)		β ($N = 148$)	
	Frequency	Fraction	Frequency	Fraction
Type of Estimates				
Aggregate	36	0.24	35	0.24
Finite Mixture	26	0.18	26	0.17
Individual Mean	64	0.43	64	0.43
Individual Median	23	0.15	23	0.16
Estimation Method				
Indifference Behavior	48	0.32	52	0.35
Logit	89	0.60	84	0.57
Probit	4	0.03	4	0.03
Other	8	0.05	8	0.05
Standard Errors				
Reported	115	0.77	110	0.74
Imputed	34	0.23	38	0.26

2011; Kerschbamer and Müller, 2020; Hedegaard, Kerschbamer, Müller and Tyran, 2021), and 6 studies recruiting a non-representative sample of adults (Dannenberg, Sturm and Vogt, 2010; Beranek, Cubitt and Gächter, 2015; Sáez, Zhu, Set, Kayser and Hsu, 2015; He and Wu, 2016; Hu, He, Zhang, Wölk, Dreher and Weber, 2018; Carpenter and Robbett, 2024). All studies offered monetary rewards for participating in the experiments.

Table 2 reports the coded features of the 297 estimates in our dataset. Around 60% of the estimates come from studies that compute individual-level estimates of α and β and then report the mean and/or the median; 18% come from four studies which use finite-mixture models (Bruhin, Fehr and Schunk, 2019; Alger and van Leeuwen, 2024; Hedegaard, Kerschbamer, Müller and Tyran, 2021; Carpenter and Robbett, 2024); and 24% come from studies which estimate parameters for a "representative" agent by pooling together all the available data. The most common econometric framework is a binary/multinomial logit model estimated by maximum likelihood, either by assumption or embedding it as a Random Utility Model with IID mean zero Gumbel noise. The second most common method is what we label "indifference threshold," where a parameter estimate is computed as the value making the subject indifferent between two outcomes. For example, a researcher can obtain an approximate estimate of α as the value making the second mover in an ultimatum game indifferent between accepting or rejecting an unfavorable offer. Finally, the parameters are elicited using choice data from a variety of games. However, even if some studies do use more complex games (e.g., sequential prisoner's dilemmas or sequential public good games), more than half of the estimates come from experiments where subjects play a combination of ultimatum games and dictator games or variations of these.

4 Results

In this section, we first provide a non-parametric description of the 149 estimates of α and 148 estimates of β in our dataset (Section 4.1). We then fit a three-level random-effects model to find average values for the advantageous and disadvantageous inequality coefficients which take into account the different degree of precision of the various estimates and the correlation between multiple estimates from the same study. This analysis, which is presented in Section 4.2, provides the main results of the paper. In addition, we try to understand the heterogeneity across studies using the features coded in our dataset (Section 4.3), and the heterogeneity at the individual level using information contained in the papers (Section 4.4). Finally, in Section 4.5, we investigate the issue of publication bias and selective reporting.

4.1 Non-Parametric Analysis

We refer back to Figure 1, presented in the Introduction, which shows the distribution of the 149 estimates of α and of the 148 estimates of β in our dataset.[11] The raw mean and median for α are, respectively, 0.35 and 0.12. In contrast with the assumption in FS ($\alpha > 0$), around a third of the estimates (53 out of 145) are equal to or less than 0.[12] This suggests

[11]Boxplots of the estimates reported in each paper can be found in Appendix J.

[12]In particular, 96 estimates are greater than 0, 8 estimates are equal to 0 and 45 estimates are smaller than 0. A z-test reveals that 82 estimates are positive and significantly (i.e., p-value < 0.05) different from 0, 35 estimates are indistinguishable from 0 and 32 estimates are negative and significantly different from 0.

Table 3: Summary Statistics for Disadvantageous Inequality (α)

	N	Min	1st Q	2nd Q	Mean	3rd Q	Max	SD
Estimate Type								
Aggregate	36	-0.14	-0.06	0.07	0.17	0.21	0.96	0.32
Finite Mixture	26	-0.09	-0.01	0.05	0.09	0.13	0.52	0.15
Individual Mean	64	-0.46	-0.02	0.31	0.56	0.98	2.81	0.75
Individual Median	23	-0.12	0.00	0.03	0.32	0.30	4.50	0.94
Experimental Task								
Game	84	-0.14	0.04	0.16	0.50	0.74	4.50	0.78
Individual Choice	65	-0.46	-0.08	0.00	0.15	0.30	1.60	0.40
Complete Dataset	149	-0.46	-0.02	0.12	0.35	0.47	4.50	0.66

that some individuals are not hurt by unfavorable comparisons with others' outcomes. Table 3 shows that the estimates of α differs depending on whether the parameter is elicited in strategic environments (i.e., situations where the decision-maker's earnings depend also on the actions of others; e.g., the ultimatum game or the prisoner's dilemma) or in individual decision-making tasks (e.g., the dictator game or choice menus).[13] In the former case, the median of α is 0.16 and in the latter case 0. This result is in line with the discussion in Dannenberg, Riechmann, Sturm and Vogt (2007), Dannenberg, Sturm and Vogt (2010), Kleine, Königstein and Rozsnyói (2014), Yang, Onderstal and Schram (2016), and He and Wu (2016) and it contributes to an ongoing debate on the economic construct captured by estimates of α. The significant difference observed in our dataset supports the hypothesis that, in strategic environments, α captures not only aversion to inequality but also other concerns, for example, reciprocity or intentions.

The estimates of β feature a bell-shaped distribution with a fatter left tail: the raw mean and median are, respectively, 0.32 and 0.30. Around a tenth of the estimates (15 out of 148) are less than 0 (in contrast with the assumption in FS).[14] This suggests that some individuals have "competitive" or "spiteful" preferences, so that they strictly prefer reducing others'

[13] The full list of games used in the 43 studies from our dataset and whether they are considered strategic environments or individual decision-making tasks can be found in Table 15 in the Appendix.

[14] In particular, 132 estimates are greater than 0, 1 estimate is equal to 0 and 15 estimates are smaller than 0. A z-test reveals that 128 estimates are positive and significantly (i.e., p-value < 0.05) different from 0, 14 estimates are indistinguishable from 0 and 6 estimates are negative and significantly different from 0.

Table 4: Summary Statistics for Advantageous Inequality (β)

	N	Min	1st Q	2nd Q	Mean	3rd Q	Max	SD
Estimate Type								
Aggregate	35	-0.46	0.24	0.33	0.33	0.48	0.76	0.24
Finite Mixture	26	-0.10	0.21	0.23	0.20	0.28	0.33	0.13
Individual Mean	64	-2.12	0.20	0.31	0.38	0.53	3.12	0.65
Individual Median	23	-0.02	0.13	0.32	0.30	0.52	0.58	0.21
Experimental Task								
Game	79	-1.27	0.19	0.23	0.31	0.32	3.12	0.48
Individual Choice	69	-2.12	0.22	0.36	0.33	0.53	0.96	0.43
Complete Dataset	148	-2.12	0.20	0.30	0.32	0.48	3.12	0.46

earnings (while keeping their own earnings unchanged). As shown in Table 4, contrary to α, estimates of β computed using choices from strategic environments seem smaller than estimates computed using choices from individual decision-making tasks, with a median of 0.23 in games compared to 0.36 in individual decision-making tasks. This difference can be due to a higher discomfort from a favorable comparison with others when the outcome is entirely attributable to one's own action and others only play a passive role. Fehr and Schmidt (2006) conjecture that the "proposer might be more fairly-mindedly in dictator games because recipients cannot stick up for themselves." Similarly, Camerer (2003) proposes a notion of responsibility such that the player who moves last and affects the other players' payoffs (as the proposer in a dictator game) feels responsible for others and treats them in a fair manner. Another potential mechanism behind greater guilt in individual decision-making tasks might be the heterogeneous effect of social image concerns: the evidence from Carpenter and Robbett (2024) suggests that the aversion to advantageous inequality measured in dictator games could be driven by the desire to follow social norms; if image concerns are stronger in non-strategic environments, perhaps because of or in combination with the notion of responsibility mentioned above, then estimates of β from individual decision-making tasks can be systematically higher than estimates from games.

Finally, we look at the joint distribution of the two parameters. Figure 4 shows a scatter plot of the 144 estimates from all studies which report estimates for both α and β. A large number of observations (98 out of 144) lie above the 45-degree line where $\alpha \leq \beta$, and this is in contrast with the behavioral assumption in FS. We can also observe three distinct patterns for the estimates: a first group with $\alpha \approx 0$ and $\beta > 0$, a second group with $\alpha > \beta > 0$ (as assumed by FS), and a third group with $\beta > \alpha > 0$. The first group includes 52 estimates from 17 studies. Most of the studies (14) use a mini dictator game with equality-efficiency trade-off to estimate the parameters. This again highlights the importance of the elicitation task for parameter estimates (and the underlying subjects' preferences). This group does not seem to differ from the others in other features of the studies or the estimates. The second group is composed of 31 estimates from 15 studies: 5 studies use a combination of games (ultimatum games) to estimate α and individual decision-making tasks (mini dictator games) to estimate β; 5 use games to estimate both parameters; and 5 use individual decision-making tasks to estimate both parameters. The last group is made of 31 estimates from 13 studies. In this case, tasks are more heterogeneous: 7 studies employ strategic environments for α, while 6 employ dictator games or variation thereof. Finally, the correlation between the two parameters is slightly positive but not significantly different from 0 ($\rho = 0.09$; $p = 0.25$). This is in line with the results discussed in Dannenberg, Riechmann, Sturm and Vogt (2007) Dannenberg, Sturm and Vogt (2010), Daruvala (2010), Blanco, Engelmann and Normann (2011), Morishima et al. (2012) and Beranek, Cubitt and Gächter (2015). This evidence suggests that the two parameters capture two separate traits of an individual's social preferences which are uncorrelated with each other or, at least, whose relationship is unclear.

4.2 Meta-Analytic Synthesis

The non-parametric analysis from Section 4.1 suffers from two potential pitfalls. First, all estimates are given equal weight, even if the information available to us suggests that the parameters computed in some studies are more reliable (i.e., more precisely estimated) than others. Second, estimates are assumed to be independent from one another, even if the

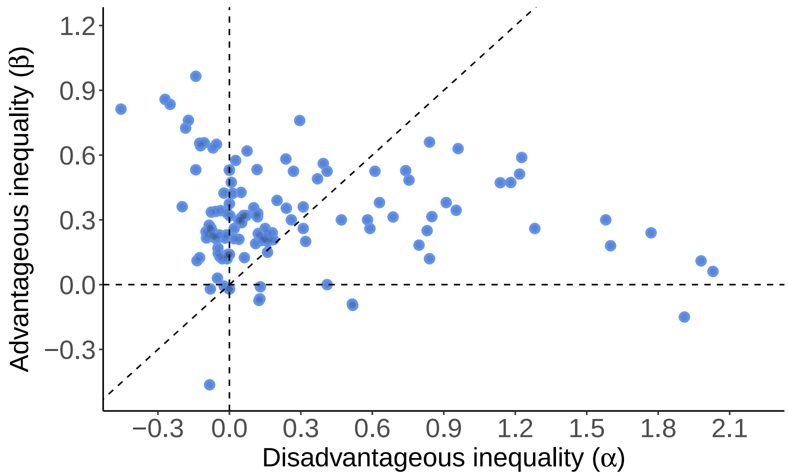


Figure 4: Scatter Plot of Disadvantageous (α) and Advantageous Inequality (β) Coefficients. Notes: We use the 144 estimates for which we have both a value for α and β; the vertical axis is truncated at -0.4 and 2.2 for better visual rendering.

same source and experimental study provides multiple estimates which are likely correlated with one another (e.g., because they are meant to capture the same subjects' underlying preferences). To tackle both these issues, we employ a three-level random effects model as in Konstantopoulos (2011) and Van den Noortgate et al. (2013), which will provide a meta-analytic estimation of a "weighted average" for α and β. From this point on, our discussion of the methodology will focus on α, considering that the same concepts and equations (up to replacing α with β) also apply to β.

Denote with α_{ij} the jth estimate of parameter α from study i. Then, the first level is defined as:

$$\alpha_{ij} = \mu_{ij} + \epsilon_{ij}, \tag{3}$$

where μ_{ij} is the "true" effect size (in this case, the "true" disadvantageous inequality parameter) and the error term represents the sampling variability, which is distributed as $\epsilon_{ij} \sim \mathcal{N}(0, v_{ij}^2)$, where v_{ij}^2 is the known sampling variance (i.e., the variance of the estimates in our dataset). The second level is:

$$\mu_{ij} = \theta_i + \xi_{ij}^{(2)}, \tag{4}$$

where θ_i represents the average disadvantageous inequality in study i and $\xi_{ij}^{(2)} \sim \mathcal{N}(0, \tau_{(2)}^2)$. The superscript (subscript) (2) in $\xi_{ij}^{(2)}$ ($\tau_{(2)}^2$) refers to the second level, which can be interpreted as the study level. The third level is:

$$\theta_i = \alpha_0 + \xi_i^{(3)}, \tag{5}$$

where α_0 is the population mean of α (what we are interested in) and $\xi_i^{(3)} \sim \mathcal{N}(0, \tau_{(3)}^2)$. Similarly, the superscript (subscript) (3) in $\xi_{ij}^{(3)}$ ($\tau_{(3)}^2$) refers to the third level, which can be interpreted as the population level. We can combine the three levels into a single equation to have:

$$\alpha_{ij} = \alpha_0 + \xi_{ij}^{(2)} + \xi_i^{(3)} + \epsilon_{ij}. \tag{6}$$

As we can see, there are two heterogeneity terms in addition to the sampling error: $\xi_{ij}^{(2)}$ represents the within-cluster heterogeneity, i.e., the heterogeneity that is present among different estimates in a single study; $\xi_i^{(3)}$, instead, stands for the between-cluster heterogeneity, with a large value for $\tau_{(3)}^2$ indicating that the "true" disadvantageous inequality parameter varies a lot between different studies. Before fitting the three-level random-effects model described above, we run some diagnostic checks to exclude potentially "overly influential" studies by computing DFBETAS (Belsley, Kuh and Welsch, 1980), which measure the effect of dropping one study on a regression coefficient. We use the classification in Bollen and Jackman (1985) and identify a study to be influential if $|\text{DFBETAS}| > 1$.

Tables 5 report the results of the meta-analytic synthesis. In discussing these results, we focus on the estimates obtained in the full sample, that is, without removing studies whose SEs we had to approximate. Results for the restricted sample of studies with reported SEs are available in the same table and are qualitatively identical. Starting with the disadvantageous inequality parameter (α), the coefficient is positive and significantly different from zero. Our meta-analysis, thus, supports the hypothesis that people are concerned about equity when they are in a disadvantageous situation. However, the value of 0.533 is

Table 5: Meta-Analytic Averages

	α		β	
	(1)	(2)	(3)	(4)
Constant	0.533	0.434	0.326	0.337
	(0.110)	(0.093)	(0.036)	(0.033)
p-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
I^2_{within}	7.80	15.96	35.29	26.08
$I^2_{between}$	92.19	84.03	64.19	72.63
Observations	149	113	144	106
Sample	Full	Restricted	Full	Restricted

Notes: All columns estimate a three-level random-effects model with the restricted maximum likelihood method; columns (2) and (4) focus on studies with reported (i.e., non-approximated) SEs; p-values are for a two-sided test with null hypothesis H_0 : Constant = 0; SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010). In column (2) the study by Diaz et al. (2023) was removed because $|DFBETAS| > 1$. In columns (3) and (4) the study by Bellemare et al. (2008) was removed because $|DFBETAS| > 1$.

smaller than the average value reported in FS (0.850). From the I^2 statistics (Higgins and Thompson, 2002), we observe that around 7.80% of the variability in the data is due to heterogeneity within studies (I^2_{within}), 92.19% to heterogeneity across studies ($I^2_{between}$), and the remainder to sampling variance. It is important to note that I^2_{within} captures within-group heterogeneity resulting solely due to error. Although heterogeneity across subjects can add to this error, other factors, including measurement error (e.g., from design choices and detectable parameter ranges) and sample size, also play a role.

The meta-analytic average of β is 0.326, and statistically different from zero at any conventional significance level. This value is in line with the weighted average of β from the distribution reported in FS (0.315). We, thus, find evidence of equity concerns in the realm of advantageous situations. The I^2 statistics shows that around 35.29% of the variability is due to within study heterogeneity and around 64.19% to between studies heterogeneity. Finally, while the theoretical assumptions in FS hold in our meta-analysis, since $\alpha \geq \beta$ and $0 \leq \beta < 1$, the estimate of α is statistically indistinguishable from the estimate of β.

In the Appendices, we offer several robustness checks. Appendix D presents the results of a Bayesian hierarchical model.[15] Appendix E offers a multivariate version of our three-level random effects model, where we jointly estimate both parameters and their correlation at the paper and population level. Appendices F.1 and F.2 provide additional robustness checks for the disadvantageous inequality parameter α, by looking at different normalizations of the parameter space and at a different assumption on estimate independence respectively. Finally, Appendix F.3 extends the three-level model by allowing estimates from the same study to have correlated estimation errors. We note that the results presented in the main body of the paper hold across all robustness checks, and that the population-level correlation in the joint model is close to zero.

4.3 Explaining Heterogeneity

The estimates in our dataset come from studies that are very different from each other, for example, because of the subject population, the tasks subjects performed during the experiment, the geographic location of the experiment and so on. It is then far fetched that the estimates for α and β depend mainly on sampling errors, either at the observation or study level, as we did previously. In order to explain the heterogeneity, we add to the three-level specification described in equation (6) a set of regressors:

$$\alpha_{ij} = \alpha_0 + \delta X_{ij} + \xi_{ij}^{(2)} + \xi_i^{(3)} + \epsilon_{ij}. \tag{7}$$

where X_{ij} is a set of moderator variables coded in our dataset. Given the high amount of coded variables and the few observations for some of these, it is unclear what model should we use to explain the heterogeneity in the parameters. We then run four different regressions, starting from a parsimonious model that only investigates the role of strategic versus non-strategic environments, and then adding other potentially relevant covariates in

[15] The main model from Appendix D assumes that the parameters are distributed normally both at the paper and at the population level. In the same Appendix, we also estimate the Bayesian hierarchical model with a different assumption on the population-level distribution of α to better fit the skewness observed in the empirical distribution.

Table 6: Explaining Heterogeneity

	Disadvantageous Inequality (α)				Advantageous Inequality (β)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.266***	0.363***	0.323***	0.496***	0.387***	0.367***	0.440***	0.452***
	(0.096)	(0.104)	(0.106)	(0.143)	(0.038)	(0.038)	(0.070)	(0.068)
Experimental Task: Strategic	0.461**	0.452**	0.489***	0.611***	-0.173**	-0.184**	-0.207**	-0.170*
	(0.192)	(0.189)	(0.146)	(0.195)	(0.074)	(0.075)	(0.101)	(0.098)
Type of Estimate: Median		-0.195*	-0.194*	-0.195*		0.007	0.007	0.005
		(0.105)	(0.105)	(0.108)		(0.033)	(0.034)	(0.033)
Type of Estimate: Aggregate		-0.183***	-0.186***	-0.141**		0.077*	0.072*	0.075
		(0.065)	(0.063)	(0.052)		(0.041)	(0.042)	(0.046)
Type of Estimate: Finite Mixture		-0.148**	-0.150**	-0.122**		0.027	0.032	0.033
		(0.059)	(0.056)	(0.050)		(0.031)	(0.034)	(0.039)
Geographic Location: North Europe			0.003	-0.096			-0.081	-0.096
			(0.148)	(0.193)			(0.094)	(0.087)
Geographic Location: USA			0.053	0.009			-0.028	-0.020
			(0.324)	(0.323)			(0.119)	(0.116)
Geographic Location: China			-0.481***	-0.462**			0.092	0.093
			(0.136)	(0.201)			(0.158)	(0.171)
Geographic Location: Multiple			-0.403**	-0.562**			0.201	0.175
			(0.189)	(0.221)			(0.157)	(0.155)
Implementation: Online			-0.295	-0.328			0.029	0.024
			(0.233)	(0.286)			(0.146)	(0.155)
Subject Population: Non Student			0.433**	0.474*			-0.168*	-0.167*
			(0.214)	(0.271)			(0.085)	(0.094)
Estimation: Logit				-0.356*				-0.028
				(0.202)				(0.069)
Estimation: Probit				0.051				0.016
				(0.109)				(0.132)
Estimation: Other				-0.399				-0.114
				(0.514)				(0.097)
I^2_{within}	8.47	8.82	8.10	8.67	38.45	40.48	42.90	41.10
$I^2_{between}$	91.52	91.17	91.89	91.32	60.97	58.93	56.48	58.31
$pseudo\text{-}R^2_{between}$	8.68	13.52	4.90	10.40	13.84	18.92	26.21	19.86
Observations	149	149	149	149	144	144	144	144

Notes: SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010). In the columns for β the study by Bellemare et al. (2008) was removed because $|DFBETAS| > 1$. *p<0.1; **p<0.05; ***p<0.01.

our dataset.

Since X_{ij} is composed of dummy variables, each coefficient represents the shift of the

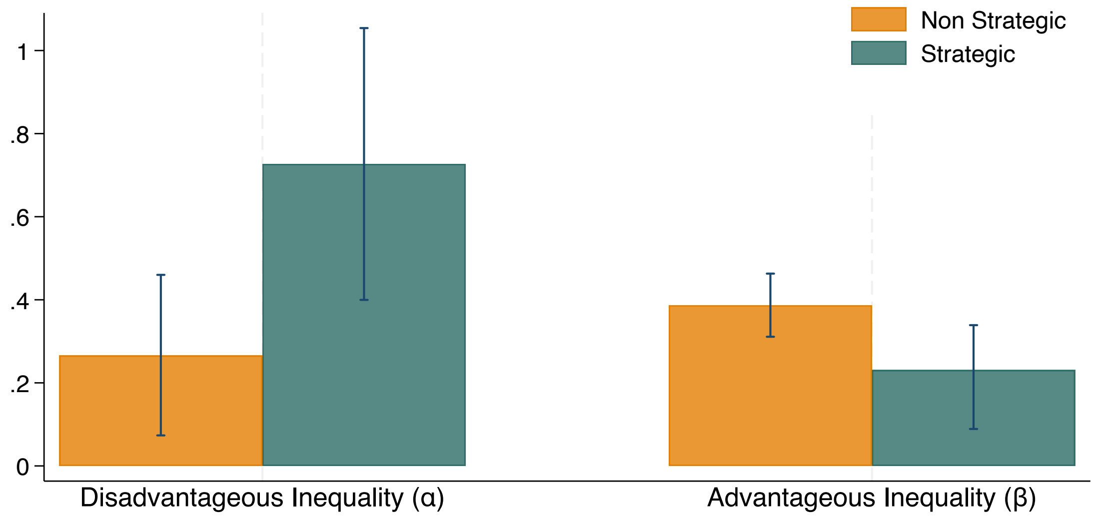


Figure 5: ML Estimates of Disadvantageous Inequality (α) and Advantageous Inequality (β) in the Sub-Samples with Strategic vs Non Strategic Environments. Notes: Estimates are from columns 1 and 5 in Table 6; vertical bars are 95% confidence intervals.

population mean α_0 with respect to the baseline condition. The meta-regressions for α and β are presented in Table 6. A positive coefficient indicates more sensitivity (i.e., stronger aversion) to disadvantageous or advantageous inequality compared to the baseline condition, and we chose the baseline conditions as follows: for the experimental task the omitted category is individual decision-making tasks; for the type of estimate, the omitted category is mean; for the subject population, the omitted category is college students (and we pool non-representative samples of adults and representative samples in the non-students category); for geographic location, the omitted category is Southern Europe; for experimental implementation, the omitted variable is in-person (where we pool laboratory and classroom experiments); and for estimation method, the omitted category is indifference threshold.[16]

While we have a small number of observations for some categories and should thus be cautious in inferring too much from these coefficients, we nonetheless highlight some inter-

[16]When the same study offers to our dataset both estimates computed in strategic environments and estimates computed in individual-decision making environments as in Yang, Onderstal and Schram 2016, we consider the two estimates as coming from two different studies. This allows for a crispier analysis of estimates' heterogeneity in this dimension, with results similar to those we obtain when estimating a three-level random-effects model on two sub-samples, one for estimates computed in strategic environments and one for estimates computed in individual-decision making environments (see Appendix G).

esting patterns. First, the population averages for α and β are strongly affected by the experimental task (see also Figure 5): sensitivity to disadvantageous inequality is stronger in strategic environments than in individual decision-making tasks, while the reverse is true for sensitivity to advantageous inequality. Therefore — since both parameters are strictly positive and, thus, capturing inequality aversion in both environments — strategic environments dampen the guilt from being ahead in social comparisons and, at the same time, they enhance the envy from being behind.[17]

Second, we learn that aggregate estimates, median estimates, and estimates from finite mixture models for α are lower than mean estimates. If the distribution of individual-level estimates of α in the population is right skewed, as we document for a subset of studies in the following subsection, then mean estimates can be systematically higher than median estimates, aggregate estimates, and estimates from finite mixture model. Third, samples composed of college students are less averse to disadvantageous inequality and more averse to advantageous inequality than other samples.[18] Fourth, estimates are uncorrelated with whether a study is conducted in-person or online. Finally, we do not see systematic differences with respect to geographic location or participants' nationality, with the exception of Chinese participants being less concerned with disadvantageous inequality than samples from Europe and North America.

To conclude, we discuss the measure of *pseudo*-$R^2_{between}$, which indicate the proportional reduction in the amount of between variance when going from the model with covariates to

[17]In Appendix H, we offer a deeper analysis of the correlation between estimates and the experimental task where we unpack strategic environments into ultimatum games and other games (a finer classification that nonetheless allows us to keep a substantial sample size in each class; individual decision-making tasks are all variations of the dictator game so we are unable to offer a finer classification). We note that 5 studies use choices from both games and individual decision-making tasks and that Alger and van Leeuwen (2024) use choices from both ultimatum games and other games. We perform the analyses both keeping these studies in the sample (and, as we do in Table 6, classifying the estimate depending on the most frequent experimental task) and excluding these 6 studies. We highlight three results from Appendix H. First, estimates of α from ultimatum games are larger than estimates from individual decision-making tasks but this is not true for estimates from other games. Second, estimates of β from both ultimatum games and other games are smaller than estimates from individual decision-making tasks. Third, while there is some evidence that estimates of α from ultimatum games are larger than estimates from other games, this difference is not robust to excluding studies that use both kinds of games and estimates of β do not depend on the class of games.

[18]The former finding is in line with the discussion in Fehr and Charness (2025).

the model without covariates. In our best-performing regressions, we explain approximately 13.5% of the variance in α and 26% of the variance in β. While our contribution here is more speculative than strictly analytical, we would like to discuss the potential reasons behind these numbers. A likely reason for the modest *pseudo*-$R^2_{between}$ values in our regressions is the widespread use of linear FS models (see Table 30, Appendix K.1), which may omit key explanatory variables. Outcome-based models account for part of the observed variation but leave many factors unexplained. These may include reciprocity, intention-based preferences, moral wiggle room, social norms, framing effects, stake size, the difference between active decision-making and passive observation, and interactions with one versus multiple agents. Although some studies in our dataset explore these factors, they are too few to draw meta-analytic conclusions.

Another interesting pattern is the difference between α and β in the *pseudo*-$R^2_{between}$. One possible explanation is that the variability of β is inherently smaller. This is already suggested by our non-parametric analysis (Table 3 and Table 4), where the standard deviation of β is consistently lower than that of α—especially in strategic environments, where α shows the greatest variability. Further evidence comes from studies in our dataset that provide individual-level estimates. In both strategic and non-strategic settings, α appears more volatile, with individuals ranging from slightly altruistic to strongly envious (i.e., large positive α). In contrast, individuals with $\beta < 0$ (i.e., spiteful preferences) are relatively rare, and most individuals exhibit only moderate levels of advantageous inequality aversion. We also conjecture that β may be less sensitive to the contextual or psychological influences outlined above, which could help explain why our models better capture its variation.

4.4 Heterogeneity in Individual-Level Estimates

Our meta-analysis takes as unit of observation either aggregate measures (that is, estimates for a "representative subject" in an experiment) or measures of central tendency (that is, means or medians for subjects participating to the same experiment) of the parameter estimates. This is the information commonly available in the papers performing the structural estimation of the FS model and, thus, using this unit of observation allows us to collect the largest number of estimates and to make the most reliable claims on what values of the parameters are most plausible. For similar reasons, it is the approach adopted by other recent meta-analyses in behavioral economics (Imai et al., 2021; Brown et al., 2024). The analysis in Section 4.3 above explores how these aggregate or central tendency measures depend on features of the study and features of the estimation methodology. Some feature of the studies have to do with the socio-demographics of their participants (for example, the location of an in-person study or, for online studies, the nationality of the participants; and whether the study is conducted with a sample composed exclusively of college students or not) and this allows us to say something about how those aggregate or central tendency measures vary across samples. At the same time, our meta-analysis does not allow us to say much about how parameter estimates vary within samples, that is, to discuss heterogeneity in individual-level estimates. In this Section, we summarize all the information available on this topic in the studies we reviewed to compose our dataset of parameter estimates.

Types from Finite Mixture Models and Clustering. Bruhin et al. (2019) conducted experiments using the Dictator Game and Reciprocity Game, identifying three types of behavior within a finite mixture model: Strongly Altruistic individuals, who form 40% of the sample and show a high willingness to increase others' payoffs regardless of whether they are ahead or behind, Moderately Altruistic individuals, constituting 50% of the sample with a lower yet positive inclination to help others, and Behindness Averse individuals, making up 10% of the sample, who are inclined to reduce others' income when they themselves are behind. Notably, purely selfish types did not emerge, and all three types valued others' payoffs more significantly when they were ahead rather than behind.

Alger and van Leeuwen (2024) used the Sequential Prisoner's Dilemma, Trust Game, and Ultimatum Game, discovering two primary types within a finite mixture model: Inequality

Averse individuals, who are aheadness and behindness averse and represent 62% of the sample, and Behindness Averse individuals, constituting 28%. They also identified a third type, Aheadness Averse, representing 17% of the sample, in a different model variation. Carpenter and Robbett (2024) employed the Dictator Game, finding three types within their finite mixture model: Strongly Altruistic individuals (52%), Moderately Altruistic individuals (43%), and a small group of Behindness Averse individuals (6%).

In addition to these three studies from our dataset, two recent studies which do not belong to our dataset use a non-parametric Bayesian clustering method (Dirichlet Process Means) to identity preference clusters.[19] Fehr et al. (2024) analyze Swiss samples and identify three preference clusters: Inequality Averse individuals (45%-53%), Altruistic individuals (30%-40%), and Predominantly Selfish individuals (10%-24%). Finally, Fehr and Charness (2025) applied the Dirichlet Process Means approach to a Danish sample from Epper et al. (2020), revealing three preference clusters: Altruistic individuals (30.2%), Inequality Averse individuals (37.3%), and Predominantly Selfish individuals (32.5%).

Distribution of Individual-Level Estimates. Tables 7 and 8 summarize the information from all papers in our dataset that report how parameter estimates are distributed in their sample. The tables show that there is significant individual level heterogeneity. At the same time, the distribution of individual level estimates into the available brackets resembles the distribution of aggregate or central tendency estimates from our dataset: the distribution of α is right-skewed — $\alpha < 0.4$ is the modal bracket in 9 studies out of 15 and at least 40% of subjects have an estimated α in this bracket in 10 studies out of 15 — and the distribution of β is instead more symmetric around the middle bracket (that is, β between 0.25 and 0.50). Moreover, the difference in these distributions can be attributed to the use of strategic versus non-strategic decision-making tasks in the experiment.

[19]These studies do not contribute observations to our dataset because they do not offer quantitative details on the individual-level estimates or on the distribution of estimates within each identified cluster, for example, the mean estimate of α and β for each cluster.

Table 7: Heterogeneity in Individual-Level Estimates, Disadvantageous Inequality (α)

Study	$\alpha < 0.50$	$0.50 \leq \alpha < 1$	$1 \leq \alpha < 4$	$\alpha \geq 4$
FS (S)	30%	30%	30%	10%
BEN (S)	31%	33%	23%	13%
BCG Nottingham (S)	54%	18%	21%	7%
BCG Izmir (S)	59%	12%	5%	24%
BCG MTurk (S)	46%	17%	20%	17%
MR (S)	42%	14%	23%	21%
HMN Random (S)	30%	37%	2%	31%
HMN Fixed (S)	31%	33%	7%	29%
DRS (NS)	90%	8%	2%	0%
DHIZ (S)	30%	10%	60%	0%
HW (NS)	90%	10%	0%	0%
YOS (NS)	98%	2%	0%	0%
CEHZ (NS)	69%	31%	0%	0%
WYSYZZJZ Gain (S)	72%	28%	0%	0%
WYSYZZJZ Loss (S)	44%	39%	17%	0%

Notes: FS: Fehr and Schmidt (1999), BEN: Blanco et al. (2011), BCG: Beranek et al. (2015), MR: Müller and Rau (2019), HMN: Huck et al. (2001), DRS: Dannenberg et al. (2007), DHIZ: Diaz et al. (2023), HW: He and Wu (2016), YOS: Yang et al. (2016), CEHZ: Corgnet et al. (2015), WYSYZZJZ: Wu et al. (2014). Studies with (S) used games while studies with (NS) used individual decision-making tasks.

Explaining Heterogeneity in Individual-Level Estimates. Finally, a limited number of studies in our dataset report correlations between individual-level parameter estimates and the subjects' individual characteristics. Bellemare et al. (2008, 2011), Beranek et al. (2015), and Daruvala (2010) investigate correlations with gender, age, education, and income. While single studies do find some statistically significant correlations, no individual characteristic is consistently associated with a greater sensitivity to advantageous or disadvantageous inequality. Corgnet et al. (2015) and Cueva et al. (2016) show that more "impulsive" individuals, that is, individuals with lower scores in the Cognitive Reflection Test are more likely to be egalitarian or have greater distributional concerns.

Table 8: Heterogeneity in Individual-Level Estimates, Advantageous Inequality (β)

Study	$\beta < 0.25$	$0.25 \leq \beta < 0.50$	$\beta \geq 0.50$
FS (S)	30%	30%	40%
BEN (NS)	29%	15%	56%
BCG Nottingham (NS)	21%	25%	54%
BCG Izmir (NS)	16%	11%	73%
BCG MTurk (NS)	20%	19%	61%
MR (NS)	25%	17%	58%
HMN Random (S)	31%	44%	25%
HMN Fixed (S)	13%	14%	73%
DRS (NS)	35%	20%	45%
DHIZ (NS)	25%	20%	55%
HW (NS)	45%	25%	30%
YOS (NS)	74%	19%	7%
CEHZ (NS)	77%		23%
BMLLM Gains (NS)	10%	17%	73%
BMLLM Losses (NS)	27%	28%	45%

Notes: FS: Fehr and Schmidt (1999), BEN: Blanco et al. (2011), BCG: Beranek et al. (2015), MR: Müller and Rau (2019), HMN: Huck et al. (2001), DRS: Dannenberg et al. (2007), DHIZ: Diaz et al. (2023), HW: He and Wu (2016), YOS: Yang et al. (2016), CEHZ: Corgnet et al. (2015), BMLLM: Boun My et al. (2018). Studies with (S) used games while studies with (NS) used individual decision-making tasks.

4.5 Identifying Selective Reporting and Publication Bias

One aspect to keep in mind when conducting a meta-analysis is the problem of selective reporting or publication bias. The main concern arises when a theory strongly predicts certain results — for example, the magnitude or significance of some statistical relationships — and the literature anchors itself towards the same findings. This causes problems when, for example, new evidence reporting "unusual" or "unconventional" results is not taken in consideration because it goes against this norm. Articles are, then, either rejected and not published in journals or simply not written to begin with (the "file-drawer" problem). Beyond biases in the publication process, there are other sources of selective reporting that

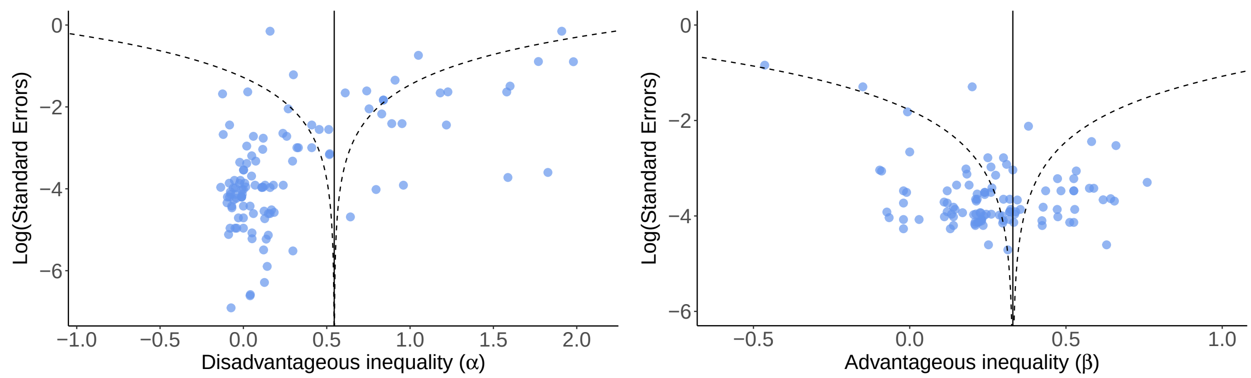


Figure 6: Funnel Plots of Disadvantageous (α) and Advantageous Inequality (β) Coefficients. Notes: The vertical continuous lines are at the meta-analytic average from columns (1) and (3) in Table 5 ($\alpha = 0.545$ and $\beta = 0.330$) and the diagonal dotted curves represent a p-value of 0.05 for a z-test whose null hypothesis is that the estimate is equal to the meta-analytic average (i.e., estimates below each dotted line are statistically different from this average). The horizontal axis is truncated at 2.2 (α); -0.6 and 1.2 (β) for better visual rendering. Only those estimates with reported SEs are included.

go from conscious frauds to unethical practices like "p-hacking".

In order to gauge the occurrence of publication bias in studies estimating other-regarding preferences parameters, we first look at funnel plots. Funnel plots are scatter plots of the parameter estimates and of their SEs. The idea is that estimates with a higher precision should lie close to the meta-synthetic mean of the parameters, while estimates far from this mean should show a lower precision. Without selective reporting, we expect to see a funnel-shaped distribution which is symmetric around the "average" parameter value. An absence of symmetry can hint to "missing" studies and so to the presence of publication bias. Figure 6 shows the funnel plots for the advantageous and disadvantageous inequality coefficients. The distribution for α looks highly asymmetric: observations with a negative (and large in magnitude) value of α which is imprecisely estimated are "missing". A similar, albeit more attenuated, effect is present also for β: there are no studies reporting a large and imprecisely estimated positive value of this coefficient.

A second approach to detect selective reporting is the FAT-PET procedure, which consists in regressing the parameters on their SEs. If there is no publication bias, the reported

Table 9: FAT-PET Analysis

	α		β	
	(1)	(2)	(3)	(4)
Constant	0.426	0.240	0.382	0.400
	(0.095)	(0.070)	(0.047)	(0.040)
Standard Errors	1.510	2.510	-1.554	-2.010
	(0.492)	(0.607)	(0.703)	(0.685)
p-value	< 0.0001	0.0018	< 0.0001	< 0.0001
I^2_{within}	8.94	22.44	35.09	24.20
$I^2_{between}$	91.05	77.54	64.39	74.53
Observations	149	113	144	106
Sample	Full	Restricted	Full	Restricted

Notes: All columns estimate a three-level random-effects model with the restricted maximum likelihood method; columns (2) and (4) focus on studies with reported (i.e., non-approximated) SEs; p-values are for a two-sided test with null hypothesis H_0 : Constant = 0; SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010); In column (2) the study by Diaz et al. (2023) was removed because $|DFBETAS| > 1$. In columns (3) and (4) the study by Bellemare et al. (2008) was removed because $|DFBETAS| > 1$.

estimates should be uncorrelated with the SEs. We then estimate the two following equations:

$$\alpha_{ij} = \alpha_0 + \delta SE_{ij} + \xi^{(2)}_{ij} + \xi^{(3)}_{i} + \epsilon_{ij}. \tag{8}$$

$$\beta_{ij} = \beta_0 + \gamma SE_{ij} + \nu^{(2)}_{ij} + \nu^{(3)}_{i} + \eta_{ij}. \tag{9}$$

In this model, δ and γ capture the degree of selective reporting bias while α_0 and β_0 represent the selection-bias-corrected value of the parameters. This exercise tests at the same time for asymmetry in the funnel plots (FAT; Egger et al. 1997; Stanley 2005; Stanley and Doucouliagos 2017) and for a "true effect" of the parameters beyond publication selection (PET). As reported in Table 9, the coefficient for δ is positive and statistically significant, while the coefficient for γ is negative and statistically significant ($\delta = 1.510$ with p-value$= 0.004$ in the full sample; $\gamma = -1.554$ with p-value$= 0.033$ in the full sample). At the same time, the constants, α_0 and β_0, are positive and highly significant (both in the full and in the restricted sample), indicating the presence of both disadvantageous and advantageous inequity aversion even after correcting for possible publication bias: the publication-bias-corrected 95% confidence intervals for α and β are, respectively, $[0.235, 0.618]$ and $[0.288, 0.476]$.

We note that the asymmetry in the funnel plots could be generated also in the absence of publication bias — for example, because of constraints in the estimation of α and β when eliciting these parameters with the experimental tasks typically employed by the literature.[20] Moreover, while the funnel plot procedure assumes that the two parameters can take any value, some values are more plausible than others since these coefficients are meant to capture social preferences. In particular, it would be surprising to find values of α smaller than -1 and values of β larger than 1, which imply that an individual is willing to burn money to increase the gap in outcomes when behind or to reduce it when ahead. Indeed, the 149 estimates of α in our dataset never take values smaller than -1 and only 4 out of 148 estimates of β take values larger than 1. The asymmetry in the plot can thus hardly be deemed sure proof of publication bias.

Another form of publication bias consists in the practice of p-hacking. Journals might be biased in publishing statistically significant results and, in turn, researchers might be tempted to push analyses just below a threshold (e.g., a p-value of 5%) by, for example, changing econometric specification or the number of covariates in a regression. Two tools employed in the literature to detect publication bias in the form of p-hacking are the histograms of z-statistics and the p-curve (Simonsohn, Nelson and Simmons, 2014). Under the presence of p-hacking, we would see a bunching of z-statistics right above the threshold of statistical significance at the 5% level, i.e., $|1.96|$. This is because researchers who obtain z-statistics just below this value have an incentive to push it right above, thus creating a discontinuity around $|1.96|$ in the histograms. From the top panels of Figure 7, we see a jump just above 1.96 in the histogram for the disadvantageous inequality parameter. While this could suggest the presence of p-hacking (with researchers pushing statistical significance above 5% to show that α is greater than zero), we must note that we have very few observations around the

[20] For example, the ultimatum and dictator games used in Blanco, Engelmann and Normann (2011) lead to feasible estimates in the following ranges: $\alpha \in [0, 4.5]$ and $\beta \in [0, 1]$.

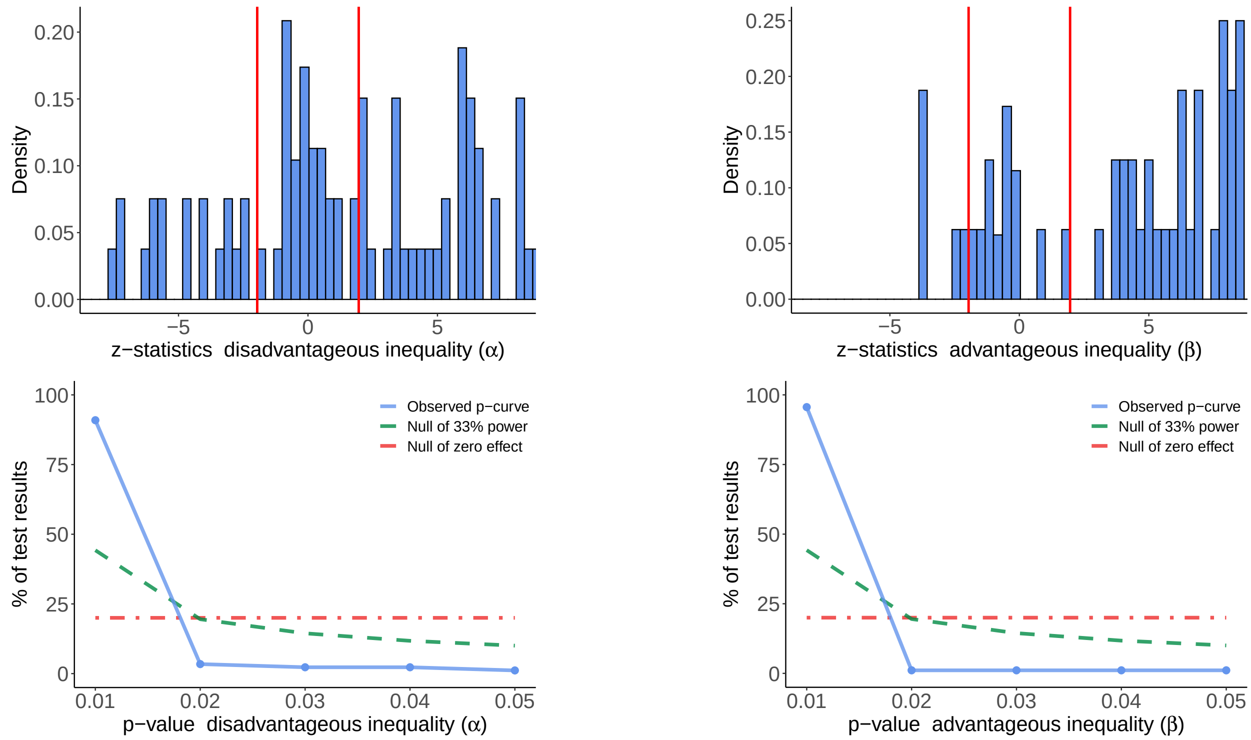


Figure 7: Distribution of z-statistics, top-panels, and p-curves, bottom-panel, of Disadvantageous inequality (α) and Advantageous Inequality (β). Notes: in both figures the test is the null hypothesis of $\alpha = 0$ or $\beta = 0$; red vertical lines are at -1.96 and 1.96. Only those estimates with reported SEs are included.

$|1.96|$ cutoff making this far from a conclusive proof.

The p-curve looks, instead, at the distribution of statistically significant p-values. Under the null hypothesis — which, in our case, is that the parameter is equal to zero — the expected distribution of statistically significant p-values is a uniform (by the definition of p-values) and we expect to see a flat p-curve. If the null hypothesis is false (that is, the parameter is different from zero) and researchers do not engage in p-hacking, we expect to see a right-skewed distribution, since researchers are more likely to find and to report small p-values rather than large ones. If the null hypothesis is true but researchers do engage in p-hacking, researchers try to turn non-significant results into significant ones and, most likely, they stop as soon as they reach this goal. In this case, we expect to see a left-skewed distribution, since researchers add to the true flat distribution of statistically significant p-values, observations that are pushed just above the 5% significance threshold. The bottom panels of Figure 7 show that the p-curves for both α and β are highly right-skewed, thus strongly supporting the hypothesis that the parameters are different from zero and the absence of p-hacking.[21]

Given the three diagnostic tools used, we conclude that there is no compelling evidence of selective reporting or publication bias. To obtain a broader perspective on these results, we conclude this section by briefly discussing the results about selective reporting and publication bias from two recent meta-analyses investigating loss aversion, Brown et al. (2024), and present bias, Imai, Rutter and Camerer (2021). The meta-analysis on loss aversion does not find compelling evidence of selective reporting and publication bias, while the meta-analysis on present bias does find modest selective reporting in real-effort tasks, in the direction of underreporting values of $\rho > 1$ (meaning "future bias"). We believe that selective reporting might be less of an issue in the Fehr and Schmidt (1999) model compared to other contexts, as the model makes sensible predictions no matter the sign and statistical significance of the parameters. As an example, finding $\alpha > 0$ implies behindness aversion, $\alpha = 0$ implies selfhishness, and $\alpha < 0$ implies behindness loving. All these results are plausible. This contrasts with other models, like the quasi-hyperbolic discounting one, where estimates of $\rho > 1$ might be viewed as unreasonable, given the anecdotal and scientific evidence pointing to present (rather than future) bias.

5 Beyond FS: Estimates of Altruistic CES Preferences

FS is not the only model of distributional preferences proposed by economists and other social scientists. Other popular models include those by Bolton and Ockenfels (2000) and Charness and Rabin (2002). We note that some important features of these models are common to the FS model and, indeed, our dataset includes studies that estimate the model

[21]While a useful instrument to detect p-hacking, the p-curve is not a definitive test. For example, if studies are well powered, the p-curve is right-skewed even in the case of a true null and mild p-hacking. Moreover, we note that some of the assumptions in (Simonsohn et al., 2014) are not satisfied in our data: many studies do not test directly the null hypothesis that the parameter is equal to zero and not all p-values come from independent studies.

by Charness and Rabin (2002).[22]

To make our systematic investigation of distributional preferences more comprehensive, we complement our earlier results with an additional meta-analysis of parameter estimates from the Altruistic Constant Elasticity of Substitution (CES) Preferences model proposed by Andreoni and Miller (2002) (AM henceforth). This analysis can provide interesting insights, as it is meant to capture altruism and attitude towards the efficiency-equity tradeoff rather than sensitivity to inequality. Moreover, this is an influential framework that has been followed by many attempts to structurally estimate its parameters with experimental data and, thus, it is the ideal candidate for a meta-analysis.

We deliberately analyze estimates from CES-based models separately from those based on inequality aversion models such as FS, and do not attempt to pool them in a unified meta-analysis as the two classes of models are conceptually and empirically distinct. Most importantly, the CES utility framework assumes that the marginal utility of the other's payoff is always non-negative, which rules out disadvantageous inequality aversion—a key feature captured by a positive α in FS. This limitation is reinforced by experimental designs that typically use a modified dictator game with a negatively sloped budget line, preventing subjects from demonstrating whether they would reduce others' payoffs at a personal cost.[23] Integrating these estimates into our main meta-analysis would therefore create a systematic downward bias in the distribution of α estimates. To avoid this conflation and to respect the theoretical and empirical differences between the two modeling approaches, we analyze the CES studies in a separate section and do not include them in the aggregate estimates of α and β reported in the previous analysis.

In this Section, we present the substantive findings of this second meta-analysis and leave the details regarding the search and data construction in Technical Appendix L.

[22]see the Technical Appendix K for details.

[23]An exception is a single session in the experiment by Andreoni and Miller (2002) which also employed a modified dictator game with a positively sloped budget line. There, around 23.5% of participants behaved in a manner compatible with inequality aversion, mainly when behind.

5.1 Experimental Task and Model

The most common elicitation task for studies in our dataset is a two-person modified dictator game, where a subject needs to choose an allocation $\pi = (\pi_s, \pi_o)$ satisfying the budget constraint $p_s\pi_s + p_o\pi_o \leq m$. Subscripts s and o stand for self and other, respectively; m represents the endowment, and p_o/p_s is the relative price of giving. Andreoni and Miller (2002) assume the following CES utility:

$$U_s(\pi_s, \pi_o) = [a\pi_s^\rho + (1-a)\pi_o^\rho]^{\frac{1}{\rho}}$$

where $a \in [0, 1]$ represents the weight of one's own payoff, with $a = 1$ implying pure selfishness. $\rho \in (-\infty, 1]$ represents the curvature of indifference curves through the elasticity $\sigma = \frac{1}{\rho - 1}$ and can be interpreted as an equality versus efficiency preference parameter. Any $0 < \rho \leq 1$ indicates distributional preference weighted towards increasing total payoffs (i.e., efficiency), as a fall in the relative price of giving lowers the expenditure for resources allocated to the other as a fraction of total expenditure. Conversely, any $\rho < 0$ indicates distributional preference weighted towards reducing payoff differences (i.e., equality). The CES specification also nests common utility functions for specific parameter's values of ρ, with Leontieff utility as $\rho \to -\infty$, Cobb-Douglas utility as $\rho \to 0$, and perfect substitutes when $\rho = 1$.

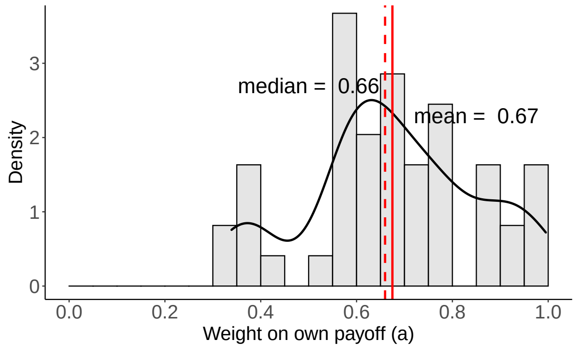
5.2 Search, Data Construction and Features of the Studies

Our inclusion criterion is: "all studies that estimate the parameters a and/or ρ from AM." The search was done on Scopus (July 3, 2023) by looking at papers that cited Andreoni and Miller (2002) and contained the stem-word "Estimat*" and the word "Elasticity." We then read all the articles that satisfied the query and applied the inclusion criterion. The final dataset comprises 18 studies from 17 papers and 98 estimates, 49 for a and 49 for ρ. The complete list of papers is available in Appendix L.1.

Table 10: Features of the Studies ($N = 18$)

	Frequency	Fraction
Publication Status		
Published	18	1.00
Methodology		
Laboratory Experiment	11	0.61
Classroom Experiment	1	0.06
Online Experiment	6	0.33
Geographic Location		
United States	12	0.66
Northern Europe (DE, NL)	4	0.22
Southern Europe (FR)	1	0.06
Africa (UG)	1	0.06
Subject Population		
College Students	12	0.66
Non-Representative Sample of Adults	6	0.34
Experimental Task		
Modified Dictator Game	15	0.83
Mini-Dictator Game with Equality-Efficiency Tradeoffs	2	0.11
Majority Bargaining	1	0.06
Reward Type		
Money	18	1.00

We then coded relevant information such as the value of the parameters, their SEs, the methodology, the subject population, the geographical location of the study (if in person) or the nationality of respondents (if online), the type of estimates (aggregate, mean, median, finite mixture), and others. Table 10 presents the features of the 18 studies in our dataset, showcasing its diversity: around two-thirds of the studies conducted laboratory experiments with college students, while a third conducted online experiments with non students; most studies, 12 out of 18, were conducted in the USA, 3 studies were conducted in Germany, 1 study each in France, Netherlands, and Uganda; the most common experimental task was the modified dictator game with a negatively sloped budget line (or an equivalent three-person version); parameters were usually estimated using a Tobit procedure on the demand function.



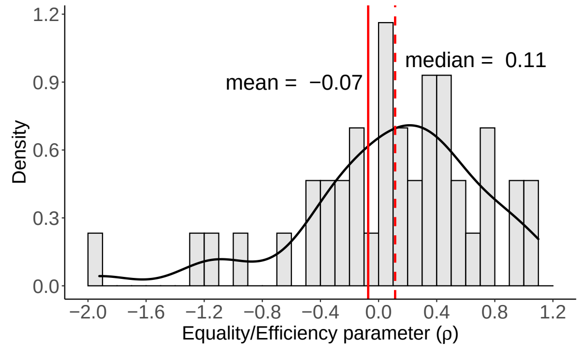


Figure 8: Distribution of the weight on own payoff (a) and equality/efficiency parameter (ρ) Coefficients. Notes: Bins for histograms are 0.05 wide for a and 0.1 for ρ; the Gaussian kernel density (solid black line) uses the Silverman's rule of thumb for bandwidth selection; in the panel for ρ, the horizontal axis is truncated at -2 and 1.2 for better visual rendering.

5.3 Results

Non-Parametric Analysis. As we can see from the histogram in Figure 8, the distribution for a is centered around a moderate level of altruism, with the mean and median equal to 0.67 and 0.66, respectively. We also see a non-negligible fraction of estimates close to 1, which indicates that some studies find selfish or close to selfish behaviour for their representative subject. The distribution for ρ lies both in the negative and the positive domain, with most estimates close to 0 or slightly positive. Studies, thus, find both negative values of ρ, indicating preferences for equality, and positive values of ρ, indicating preferences for efficiency. The correlation between the two parameters is negative, at -0.12, but not significant (p-value 0.41).

Meta-Analytic Synthesis: We now present the three-level random effects model's results, which follow the exact econometric specification in equation (6). The meta-analytic average for a is 0.686 and significantly different from 1, pure selfishness, indicating moderate levels of altruism, with agents caring around twice as much about themselves compared to the other individual. Most of the variability in the model is due to heterogeneity between papers, as seen by the $I^2_{between}$, which is close to 75% in the full sample and 90% in the restricted sample. The meta-analytic average for ρ is -0.196 but statistically indistinguishable from zero. In

Table 11: Meta-Analytic Averages

	a		ρ	
	(1)	(2)	(3)	(4)
Constant	0.686	0.688	-0.196	-0.170
	(0.038)	(0.041)	(0.215)	(0.227)
p-value	< 0.0001	< 0.0001	0.375	0.464
I^2_{within}	24.64	12.35	46.24	45.57
$I^2_{between}$	75.31	87.60	53.76	54.43
Observations	49	46	49	46
Sample	Full	Restricted	Full	Restricted

Notes: All columns estimate a multi-level random-effects model with the restricted maximum likelihood method; columns (2) and (4) focus on studies with reported (i.e., non-approximated) SEs; p-values are for a two-sided test with null hypothesis H_0 : Constant = 1 for a and Constant = 0 for ρ; SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010).

this case, we observe more importance of the within-paper heterogeneity, which corresponds to around 45% of the total variability in the model as seen by the I^2_{within}. We note that testing whether $\rho = 0$ does not reflect a clear turning point in preferences in the same way it does for the FS parameters (e.g., $\alpha > 0$ indicating envy, $\alpha < 0$ indicating altruism). Instead, ρ lies on a continuum—from perfect substitutes ($\rho = 1$) to Leontief preferences ($\rho = -\infty$)—and lacks a sharp interpretive cutoff. Still, we report tests for $\rho = 0$ for two reasons. First, they help quantify estimation uncertainty through confidence intervals together with the SEs. Second, $\rho = 0$ serves as a conventional reference in the literature, where positive (negative) values are associated with efficiency- (equality-) focused preferences (Fisman, Kariv and Markovits, 2007).

Overall, the representative agent that emerges from the meta-analytic synthesis has approximately a Cobb-Douglas utility, with weight 2/3 on one's own payoff and 1/3 on the other's payoff. In the context of the modified dictator game commonly used to estimate the parameters, this implies that the expenditure on resources allocated to self is a constant fraction a of the endowment m.

The parameter estimates for the AM model cannot be directly compared to the parameter estimates for the FS model: in AM, there is no distinction between the parameters when behind versus ahead, and the tasks used for estimation do not allow for measurement of envy or spite. Keeping this in mind, we can nonetheless make a heuristic comparison if we assume $\rho = 1$. In this case, the CES utility function from AM becomes a simple weighted average $a\pi_s + (1 - a)\pi_o$. Similarly, we can rewrite the utility function from FS to obtain a weighted average of one's own earnings and other's earnings, with weight on one's own earnings equal to $(1 - \beta)$ when ahead and $(1 + \alpha)$ when behind.[24] Given our meta-analytic average for β, 0.326, we would obtain an estimate for a equal to 0.674, which is very similar to the estimated a from Table 11. This comparison works well with β, as the CES utility and tasks can measure $\beta > 0$, and spite is uncommon in the FS dataset. This comparison is different for α, as the Altruistic CES Preferences model cannot capture $\alpha > 0$, which represents the majority of estimates in the FS dataset.

Explaining Heterogeneity: As we did for FS, we investigate whether the heterogeneity in estimated parameters can be explained by features of the estimates and of the studies. Table 12 presents the meta-regressions for a and ρ. A positive value of coefficients in columns 1-4 indicates more selfishness (with a negative value indicating more generosity); a positive value of coefficient in columns 5-8 indicates stronger preferences toward efficiency (with a negative value indicating stronger preferences toward equality). The baseline conditions are as follows: for the type of estimate, the omitted category is mean; for the subject population, the omitted category is college students; for geographic location, the omitted category is USA; for experimental implementation, the omitted variable is in-person; and for estimation method, the omitted category is Tobit. Keeping in mind the caveats discussed in Section 4.3, there are some interesting patterns. First, participants to online studies care more about equity than participants to in-person studies. Second, samples composed exclusively of college students care more about efficiency than other samples. Third, participants from

[24]When the decision-maker is ahead, the FS utility function is $\pi_s - \beta(\pi_s - \pi_o) = (1 - \beta)\pi_s + \beta\pi_o$. When the decision-maker is behind, the FS utility function is $\pi_s - \alpha(\pi_o - \pi_s) = (1 + \alpha)\pi_s - \alpha\pi_o$.

Table 12: Explaining Heterogeneity

	Weight on own payoff (a)				Equality/Efficiency (ρ)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.686***	0.720***	0.819***	0.887***	-0.196	-0.684*	-0.258	-0.212
	(0.038)	(0.034)	(0.064)	(0.043)	(0.215)	(0.390)	(0.171)	(0.217)
Type of Estimate: Median		0.000	-0.006	0.003		0.997**	1.024**	1.027**
		(0.024)	(0.024)	(0.026)		(0.351)	(0.370)	(0.382)
Type of Estimate: Aggregate		-0.143**	-0.172***	-0.164***		0.573	0.269	0.152
		(0.054)	(0.053)	(0.038)		(0.428)	(0.441)	(0.381)
Type of Estimate: Finite Mixture		-0.046	-0.062	-0.081**		0.723**	0.421	0.366
		(0.047)	(0.046)	(0.035)		(0.337)	(0.269)	(0.281)
Geographic Location: North Europe			-0.154*	-0.066*			0.207	0.111
			(0.076)	(0.034)			(0.274)	(0.273)
Geographic Location: South Europe			-0.246***	-0.019			1.809***	1.873***
			(0.064)	(0.013)			(0.171)	(0.191)
Geographic Location: Africa			-0.002	0.233***			0.309	0.468
			(0.084)	(0.070)			(0.390)	(0.389)
Implementation: Online			-0.113	-0.186***			-0.588***	-0.635***
			(0.072)	(0.046)			(0.116)	(0.180)
Subject Population: Non Student			0.010	-0.006			-0.985***	-0.961***
			(0.058)	(0.058)			(0.139)	(0.137)
Estimation: Logit				-0.311***				0.227
				(0.026)				(0.553)
Estimation: Other				-0.295***				-0.110
				(0.046)				(0.180)
I^2_{within}	24.64	28.26	29.78	63.50	46.24	27.92	77.76	73.71
$I^2_{between}$	75.31	71.69	70.16	36.38	53.76	72.08	22.23	26.29
$pseudo\text{-}R^2_{between}$	0.00	13.49	19.03	80.41	0.00	0.00	85.14	80.89
Observations	49	49	49	49	49	49	49	49

Notes: SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010). *p<0.1; **p<0.05; ***p<0.01.

Southern Europe care more about efficiency than participants from the USA.[25] Finally, participants from Northern Europe are more altruistic than participants from the USA.

These meta-regressions help to explain heterogeneity in aggregate estimates or central tendency measures of individual-level estimates (which are our unit of observation) but cannot speak to heterogeneity in individual-level estimates, which, as in FS, plays an important

[25]Note, however, that our database features a single study from Southern Europe.

role (Fisman, Kariv and Markovits, 2007; Fisman, Jakiela, Kariv and Markovits, 2015). For a concise description of individual-level heterogeneity in estimates of parameters from AM, we refer the reader to the recent review by Fehr and Charness (2025).[26] Fisman et al. (2017, 2023) estimate the AM preference parameters for two large and diverse samples of Americans — that is, two subsamples of subjects drawn from the American Life Panel in 2013 and in 2016. Their replication package allows us to complement the distribution of point estimates presented by Fehr and Charness (2025) with information on the SEs of these estimates (and, thus, whether these individual-level estimates are statistically distinguishable from 0 or 1). The median point estimates of a are 0.62 (2013, $N = 993$) and 0.59 (2016, $N = 687$); around 20% of participants have an estimated a indistinguishable from 1 (that is, only cares about one's own earnings); around 80% have an estimated a strictly between 0 and 1; and around 1% have an estimated a indistinguishable from 0 (that is, only care about others' earnings). The median point estimates of ρ are -0.18 (2013) and -0.09 (2016); between 17% and 20% of participants have a negative and statistically significant ρ; between 46% and 59% have an estimated ρ which is indistinguishable from 0; and between 25% and 34% have a positive and statistically significant ρ. Thus, while there is significant individual-level heterogeneity in parameter estimates, the meta-analytic "weighted average of averages" is coherent with the distribution of individual-level parameter estimates from these two populations.

Selective Reporting and Publication Bias: Finally, we look for selective reporting and publication bias. The funnel plot for a does not show evidence of selective reporting, which is also confirmed by the results of the FAT-PET Analysis in Table 13. The funnel plot for ρ exhibits an asymmetry, with estimates favouring large negative values but not large positive ones. This observation is captured by the FAT-PET analysis, particularly in the restricted sample. The results of the FAT-PET for a align with the table on the meta-analytic averages. The weight on own payoff corrected by publication bias is 0.638 and statistically different from 1 (pure selfishness) in both the full and restricted sample, while the dummy

[26]See the discussion in Appendix 1 and Table A1.

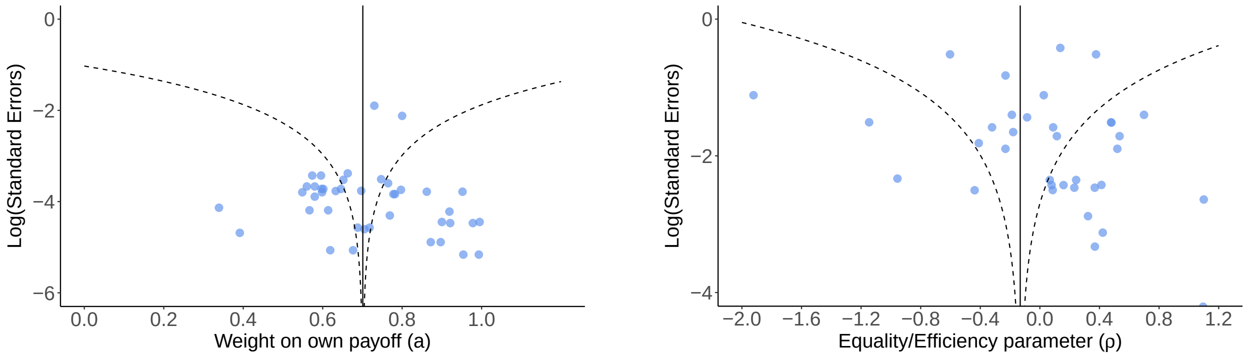


Figure 9: Funnel Plots of the weight on own payoff parameter (a) and equality/efficiency parameter (ρ) Coefficients. Notes: The vertical continuous lines are at the meta-analytic average from columns (1) and (3) in Table 11 ($a = 0.686$ and $\rho = -0.196$) and the diagonal dotted curves represent a p-value of 0.05 for a z-test whose null hypothesis is that the estimate is equal to the meta-analytic average (i.e., estimates below each dotted line are statistically different from this average). In the panel for ρ, the horizontal axis is truncated at -2 and 1.2 for better visual rendering. Only those estimates with reported SEs are included.

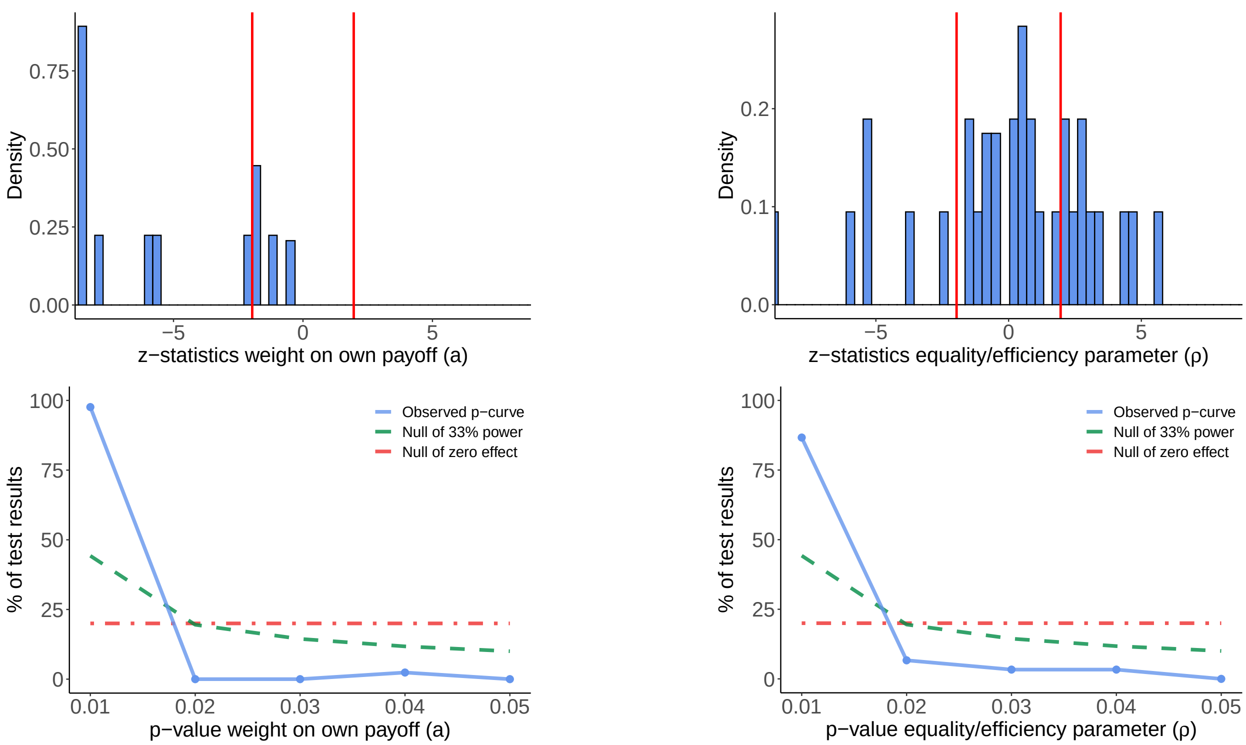


Figure 10: Distribution of z-statistics, top-panels, and p-curves, bottom-panel, of the weight on own payoff parameter (a) and equality/efficiency parameter (ρ) Coefficients. Notes: in the left figure the test is the null hypothesis of $a = 1$, in the right figure the test is the null hypothesis of $\rho = 0$; red vertical lines are at -1.96 and 1.96. Only those estimates with reported SEs are included.

Table 13: FAT-PET Analysis

	a		ρ	
	(1)	(2)	(3)	(4)
Constant	0.638	0.678	0.141	0.404
	(0.071)	(0.077)	(0.281)	(0.226)
Standard Errors	0.236	0.490	-1.621	-2.790
	(2.400)	(2.264)	(1.001)	(0.937)
p-value	< 0.0001	< 0.0001	0.622	0.093
I^2_{within}	20.09	12.18	58.45	61.34
$I^2_{between}$	79.86	87.78	41.55	38.66
Observations	49	46	49	46
Sample	Full	Restricted	Full	Restricted

Notes: All columns estimate a multi-level random-effects model with the restricted maximum likelihood method; columns (2) and (4) focus on studies with reported (i.e., non-approximated) SEs; p-values are for a two-sided test with null hypothesis H_0 : Constant = 1 for a and Constant = 0 for ρ; SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010).

on the standard errors is not significant. Considering now the equality/efficiency parameter, in column (3) both coefficients are not statistically different from 0, thus supporting, on average, Cobb-Douglas preferences and no evidence of publication bias. Once we look at the restricted sample though, ρ is greater than zero at around 10% significance level, while the SEs coefficient is negative and significant. However, this asymmetry could be attributed to the nature of the parameter itself, given the range between $-\infty$ and 1, and not necessarily by selective reporting. To conclude our comprehensive battery of rigorous statistical tests, both the p-curve and histograms of z-scores support no p-hacking. Overall, considering the robustness of these tests, there is not strong support for selective reporting or p-hacking for the estimates of the Andreoni and Miller (2002) model.

6 Conclusion

In this paper, we reported the results of meta-analyses of empirical estimates of outcome-based other-regarding preferences à la Fehr and Schmidt (1999) and à la Andreoni and Miller

(2002). We use three-level random-effects models to provide a "weighted average" for sensitivity to disadvantageous inequality (α in FS), sensitivity to advantageous inequality (β in FS), altruism (a in AM), and attitude towards equity versus efficiency (ρ in AM). We learn that the mean sensitivity to disadvantageous inequality is 0.533 with a 95% confidence interval of $[0.311, 0.755]$; the mean sensitivity to advantageous inequality coefficient is, instead, 0.326 with a 95% confidence interval $[0.254, 0.398]$. This means that, on average, individuals feel *guilt* and are willing to pay \$0.48 to increase others' earnings by \$1 when ahead; and that they feel *envy* and are willing to pay \$0.35 to decrease others' earnings by \$1 when behind.[27] The theoretical assumptions originally made in FS — that is, $\alpha \geq \beta$ and $0 < \beta < 1$ — are upheld in our empirical analysis, but we cannot conclude that the disadvantageous inequality coefficient is statistically greater than the coefficient for advantageous inequality. We also observe no correlation between the two parameters in our dataset. With respect to AM, we show that the average individual has Cobb-Douglas preferences over own and others' earnings with weight to others' earnings equal to 1/3.

An important implication of our findings is that estimated distributional preference parameters vary systematically across contexts: in particular, we observe higher values of α and lower values of β in strategic environments compared to non-strategic tasks. This pattern is consistent with the idea that the parameters of outcome-based models, such as Fehr and Schmidt (1999), may capture not only distributional concerns but also other psychological or motivational forces—such as reciprocity or concerns about intentions—that are not explicitly modeled in these frameworks. For example, responder behavior in the ultimatum game likely reflects not only aversion to disadvantageous inequality, but also negative reciprocity, i.e., a desire to punish unfair intentions—yet the estimated α parameter necessarily conflates these motives within the structure of the model.

In this sense, α and β should be interpreted as reduced-form representations of behavior that may bundle together distinct underlying drivers. As such, purely distributional models

[27] These WTPs are computed as $\beta/(1-\beta)$ when ahead and $\alpha/(1+\alpha)$ when behind.

remain valuable for their tractability and empirical relevance, but it is important to be clear about their interpretational scope. Our results help clarify this scope by showing where and how these parameters may shift and thus provide guidance for applied work using these models to explain or predict behavior in different classes of environments. For instance, applying an estimate derived from ultimatum game data to explain decisions in non-strategic tasks may lead to misleading conclusions if the role of reciprocity in the original context is not taken into account.

We also note that our analysis takes as the unit of observation either aggregate measures (that is, estimates for a "representative subject" in an experiment) or measures of central tendency (that is, means or medians for subjects participating to the same experiment) of the parameter estimates and, as such, it neglects individual-level heterogeneity. However, this simplification offers practical advantages, as incorporating individual-level or mixtures of preference types can substantially increase model complexity.

In Appendix I, we compare the predictive power of our meta-analytic estimates with the four-type and two-type mixture distributions proposed by Fehr and Schmidt (1999) and Fehr, Klein and Schmidt (2007). In this out-of-sample prediction exercise, our estimates achieve a better fit than the mixture models. These results confirm that our meta-analytic averages provide empirically disciplined and externally valid benchmarks for applied modelling.

In summary, our meta-analytic synthesis yields representative agent estimates of $\alpha = 0.533$ and $\beta = 0.326$. In practical terms, these values imply a willingness to pay of \$0.35 to reduce another's income by \$1 when behind, and \$0.48 to increase another's income by \$1 when ahead. These values provide simple, ready-to-use parameters for researchers seeking to incorporate distributional preferences into their models. Finally, our analysis suggests an important avenue for further research: studying outcome-based social preferences (e.g., inequality aversion), intention-based social preferences (e.g., reciprocity), and image concerns in the same theoretical framework and designing experiments which allow the joint estimation of parameters from these models.

References

Alger, Ingela and Boris van Leeuwen, "Estimating Social Preferences and Kantian Morality in Strategic Interactions," *Journal of Political Economy Microeconomics*, 2024, *2* (4), 665–706.

Andreoni, James and John Miller, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 2002, *70* (2), 737–753.

Anthoff, David, Cameron Hepburn, and Richard SJ Tol, "Equity Weighting and the Marginal Damage Costs of Climate Change," *Ecological Economics*, 2009, *68* (3), 836–849.

Aronsson, Thomas and Olof Johansson-Stenman, "Optimal Taxation and Other-Regarding Preferences," 2023. Unpublished Manuscript, available at SSRN 4606987.

Azar, Christian and Thomas Sterner, "Discounting and Distributional Considerations in the Context of Global Warming," *Ecological Economics*, 1996, *19* (2), 169–184.

Baranski, Andrzej and Rebecca Morton, "The Determinants of Multilateral Bargaining: A Comprehensive Analysis of Baron and Ferejohn Majoritarian Bargaining Experiments," *Experimental Economics*, 2021, pp. 1–30.

Bellemare, Charles, Sabine Kröger, and Arthur van Soest, "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," *Econometrica*, 2008, *76* (4), 815–839.

_ , _ , and _ , "Preferences, Intentions, and Expectation Violations: A Large-Scale Experiment with a Representative Subject Pool," *Journal of Economic Behavior & Organization*, 2011, *78* (3), 349–365.

Belsley, David A, Edwin Kuh, and Roy E Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Hoboken, NJ: Wiley, 1980.

Beranek, Benjamin, Robin Cubitt, and Simon Gächter, "Stated and Revealed Inequality Aversion in Three Subject Pools," *Journal of the Economic Science Association*, 2015, *1* (1), 43–58.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social

History," *Games and Economic Behavior*, 1995, *10* (1), 122–142.

Berkey, CS, DC Hoaglin, A Antczak-Bouckoms, F Mosteller, and GA Colditz, "Meta-Analysis of Multiple Outcomes by Regression with Random Effects," *Statistics in Medicine*, 1998, *17* (22), 2537–2550.

Bewley, Truman F, *Why wages don't fall during a recession*, Harvard university press, 1999.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann, "A Within-Subject Analysis of Other-Regarding Preferences," *Games and Economic Behavior*, 2011, *72* (2), 321–338.

Bollen, Kenneth A and Robert W Jackman, "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases," *Sociological Methods & Research*, 1985, *13* (4), 510–542.

Bolton, Gary E and Axel Ockenfels, "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 2000, *90* (1), 166–193.

Breza, Emily, Supreet Kaur, and Nandita Krishnaswamy, *Scabs: The social suppression of labor supply*, National Bureau of Economic Research, 2019.

_ , _ , and Yogita Shamdasani, "The morale effects of pay inequality," *The Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Brown, Alexander L, Taisuke Imai, Ferdinand Vieider, and Colin F Camerer, "Meta-Analysis of Empirical Estimates of Loss-Aversion," 2024. forthcoming Journal of Economic Literature.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk, "The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences," *Journal of the European Economic Association*, 2019, *17* (4), 1025–1069.

Burks, Stephen V, Jeffrey P Carpenter, and Eric Verhoogen, "Playing Both Roles in the Trust Game," *Journal of Economic Behavior & Organization*, 2003, *51* (2), 195–216.

Camerer, Colin F, *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, 2003.

Cantoni, Davide, Louis-Jonas Heizlsperger, David Y Yang, Noam Yuchtman, and Y Jane Zhang, "The fundamental determinants of protest participation: Evidence from Hong Kong's antiauthoritarian movement," *Journal of Public Economics*, 2022, *211*, 104667.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell, "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 2017, *76* (1).

Carpenter, Jeffrey and Andrea Robbett, "Measuring Socially Appropriate Social Preferences," *Games and Economic Behavior*, 2024, *147*, 517–532.

Charness, Gary and Matthew Rabin, "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, 2002, *117* (3), 817–869.

Chen, Yan and Sherry Xin Li, "Group Identity and Social Preferences," *American Economic Review*, 2009, *99* (1), 431–457.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider, "Social comparison and effort provision: Evidence from a field experiment," *Journal of the European Economic Association*, 2014, *12* (4), 877–898.

Cooper, David J and E Glenn Dutcher, "The Dynamics of Responder Behavior in Ultimatum Games: a Meta-Study," *Experimental Economics*, 2011, *14* (4), 519–546.

_ **and John H Kagel**, *Other Regarding Preferences: A Selective Survey of Experimental Results. The Handbook of Experimental Economics, Volume 2*, Princeton University Press, 2016.

Corgnet, Brice, Antonio M Espín, and Roberto Hernán-González, "The Cognitive Basis of Social Behavior: Cognitive Reflection Overrides Antisocial but not always Prosocial Motives," *Frontiers in Behavioral Neuroscience*, 2015, *9*, 287.

Cueva, Carlos, Iñigo Iturbe-Ormaetxe, Esther Mata-Pérez, Giovanni Ponti, Mar-

cello Sartarelli, Haihan Yu, and Vita Zhukova, "Cognitive (Ir)Reflection: New Experimental Evidence," *Journal of Behavioral and Experimental Economics*, 2016, *64*, 81–93.

Dannenberg, Astrid, Bodo Sturm, and Carsten Vogt, "Do Equity Preferences Matter for Climate Negotiators? An Experimental Investigation," *Environmental and Resource Economics*, 2010, *47* (1), 91–109.

_ , Thomas Riechmann, Bodo Sturm, and Carsten Vogt, "Inequity Aversion and Individual Behavior in Public Good Games: An Experimental Investigation," 2007. Unpublished Manuscript, available at https://www.econstor.eu/bitstream/10419/24598/1/dp07034.pdf.

Daruvala, Dinky, "Would the Right Social Preference Model Please Stand Up!," *Journal of Economic Behavior & Organization*, 2010, *73* (2), 199–208.

den Noortgate, Wim Van, José Antonio López-López, Fulgencio Marín-Martínez, and Julio Sánchez-Meca, "Three-Level Meta-Analysis of Dependent Effect Sizes," *Behavior Research Methods*, 2013, *45* (2), 576–594.

Diaz, Lina, Daniel Houser, John Ifcher, and Homa Zarghamee, "Estimating Social Preferences Using Stated Satisfaction: Novel Support for Inequity Aversion," *European Economic Review*, 2023, *155*, 104436.

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard, "Fairness and frictions: The impact of unequal raises on quit behavior," *American Economic Review*, 2019, *109* (2), 620–663.

Eckel, Catherine and Herbert Gintis, "Blaming the Messenger: Notes on the Current State of Experimental Economics," *Journal of Economic Behavior & Organization*, 2010, *73* (1), 109–119.

Eckel, Catherine C and Philip J Grossman, "Chivalry and Solidarity in Ultimatum Games," *Economic Inquiry*, 2001, *39* (2), 171–188.

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

Embrey, Matthew, Guillaume R Fréchette, and Sevgi Yuksel, "Cooperation in the Finitely Repeated Prisoner's Dilemma," *The Quarterly Journal of Economics*, 2018, *133* (1), 509–551.

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

Epper, Thomas, Ernst Fehr, and Julien Senn, "Other-Regarding Preferences and Redistributive Politics," Technical Report, Working Paper 2020.

Epper, Thomas F, Ernst Fehr, Claus Thustrup Kreiner, Søren Leth-Petersen, Isabel Skak Olufsen, and Peer Ebbesen Skov, "Inequality aversion predicts support for public and private redistribution," *Proceedings of the National Academy of Sciences*, 2024, *121* (39), e2401445121.

Fehr, Ernst, Alexander Klein, and Klaus M Schmidt, "Fairness and Contract Design," *Econometrica*, 2007, *75* (1), 121–154.

_ **and Gary Charness**, "Social Preferences: Fundamental Characteristics and Economic Consequences," *Journal of Economic Literature*, 2025, *63* (2), 440–514.

_ **and Klaus M Schmidt**, "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 1999, *114* (3), 817–868.

_ **and** _ , "Fairness and Incentives in a Multi-Task Principal-Agent Model," *The Scandinavian Journal of Economics*, 2004, *106* (3), 453–474.

_ **and** _ , "The Economics of Fairness, Reciprocity and Altruism–Experimental Evidence and New Theories," *Handbook of the Economics of Giving, Altruism and Reciprocity*, 2006, *1*, 615–691.

_ **, Michael Naef, and Klaus M Schmidt**, "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment," *American Economic Review*, 2006, *96* (5), 1912–1917.

— , **Susanne Kremhelmer, and Klaus M Schmidt**, "Fairness and the Optimal Allocation of Ownership Rights," *The Economic Journal*, 2008, *118* (531), 1262–1284.

— , **Thomas Epper, and Julien Senn**, "Social preferences and redistributive politics," *Review of Economics and Statistics*, 2024, pp. 1–45.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv, "Distributional Preferences and Political Behavior," *Journal of Public Economics*, 2017, *155*, 1–10.

— , — , — , **and Daniel Markovits**, "The Distributional Preferences of an Elite," *Science*, 2015, *349* (6254), aab0096.

— , — , — , **and Silvia Vannutelli**, "The Distributional Preferences of Americans, 2013–2016," *Experimental Economics*, 2023, *26* (4), 727–748.

— , **Shachar Kariv, and Daniel Markovits**, "Individual Preferences for Giving," *American Economic Review*, 2007, *97* (5), 1858–1876.

Forsythe, Robert, Joel L Horowitz, Nathan E Savin, and Martin Sefton, "Fairness in simple Bargaining Experiments," *Games and Economic Behavior*, 1994, *6* (3), 347–369.

Gelman, Andrew and Iain Pardoe, "Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models," *Technometrics*, 2006, *48* (2), 241–251.

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

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior & Organization*, 1982, *3* (4), 367–388.

He, Haoran and Keyu Wu, "Choice Set, Relative Income, and Inequity Aversion: an Experimental Investigation," *Journal of Economic Psychology*, 2016, *54*, 177–193.

Hedegaard, Morten, Rudolf Kerschbamer, Daniel Müller, and Jean-Robert Tyran, "Distributional Preferences Explain Individual Behavior Across Games and Time," *Games and Economic Behavior*, 2021, *128*, 231–255.

Hedges, Larry V, Elizabeth Tipton, and Matthew C Johnson, "Robust Variance

Estimation in Meta-Regression with Dependent Effect Size Estimates," *Research Synthesis Methods*, 2010, *1* (1), 39–65.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger, Natalie Smith Henrich, Kim Hill, Francisco Gil-White, Michael Gurven, Frank W. Marlowe, John Q Patton, and David Tracer, ""Economic Man" in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies," *Behavioral and Brain Sciences*, 2005, *28* (6), 795–815.

Higgins, Julian PT and Simon G Thompson, "Quantifying Heterogeneity in a Meta-Analysis," *Statistics in Medicine*, 2002, *21* (11), 1539–1558.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith, "Preferences, Property Rights, and Anonymity in Bargaining Games," *Games and Economic Behavior*, 1994, *7* (3), 346–380.

Hu, Yang, Lisheng He, Lei Zhang, Thorben Wölk, Jean-Claude Dreher, and Bernd Weber, "Spreading Inequality: Neural Computations Underlying Paying-it-forward Reciprocity," *Social Cognitive and Affective Neuroscience*, 2018, *13* (6), 578–589.

Huck, Steffen, Wieland Müller, and Hans-Theo Normann, "Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets," *The Economic Journal*, 2001, *111* (474), 749–765.

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

Ishak, K Jack, Robert W Platt, Lawrence Joseph, and James A Hanley, "Impact of Approximating or Ignoring Within-Study Covariances in Multivariate Meta-Analyses," *Statistics in Medicine*, 2008, *27* (5), 670–686.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Kaur, Supreet, "Nominal wage rigidity in village labor markets," *American Economic Review*, 2019, *109* (10), 3585–3616.

Kerschbamer, Rudolf and Daniel Müller, "Social preferences and political attitudes: An online experiment on a large heterogeneous sample," *Journal of Public Economics*, 2020, *182*, 104076.

Kirkham, Jamie J, Richard D Riley, and Paula R Williamson, "A Multivariate Meta-Analysis Approach for Reducing the Impact of Outcome Reporting Bias in Systematic Reviews," *Statistics in Medicine*, 2012, *31* (20), 2179–2195.

Kleine, Fabian, Manfred Königstein, and Balázs Rozsnyói, "Voluntary Leadership in an Experimental Trust Game," *Journal of Economic Behavior & Organization*, 2014, *108*, 442–452.

Konstantopoulos, Spyros, "Fixed Effects and Variance Components Estimation in Three-Level Meta-Analysis," *Research Synthesis Methods*, 2011, *2* (1), 61–76.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe, "Do wage cuts damage work morale? Evidence from a natural field experiment," *Journal of the European Economic Association*, 2013, *11* (4), 853–870.

Le, Minh and Alejandro Saporiti, "Inequity Aversion and the Stability of Majority Rule," 2024. Unpublished Manuscript, available at SSRN 4811942.

Ledyard, John O, "Public Goods: A Survey of Experimental Research," in John Kagel and Roth Alvin, eds., *Handbook of Experimental Economics*, Princeton University Press, 1995.

Loewenstein, George F, Leigh Thompson, and Max H Bazerman, "Social Utility and Decision Making in Interpersonal Contexts.," *Journal of Personality and Social psychology*, 1989, *57* (3), 426.

Lü, Xiaobo, Kenneth Scheve, and Matthew J Slaughter, "Inequity Aversion and the International Distribution of Trade Protection," *American Journal of Political Science*, 2012, *56* (3), 638–654.

Meager, Rachael, "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics*, 2019, *11* (1), 57–91.

_ , "Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature," *American Economic Review*, 2022, *112* (6), 1818–47.

Morishima, Yosuke, Daniel Schunk, Adrian Bruhin, Christian C Ruff, and Ernst Fehr, "Linking Brain Structure and Activation in Temporoparietal Junction to Explain the Neurobiology of Human Altruism," *Neuron*, 2012, *75* (1), 73–79.

Müller, Stephan and Holger A Rau, "Decisions Under Uncertainty in Social Contexts," *Games and Economic Behavior*, 2019, *116*, 73–95.

My, Kene Boun, Nicolas Lampach, Mathieu Lefebvre, and Jacopo Magnani, "Effects of Gain-Loss Frames on Advantageous Inequality Aversion," *Journal of the Economic Science Association*, 2018, *4* (2), 99–109.

Normann, Hans-Theo and Holger A Rau, "Simultaneous and Sequential Contributions to Step-Level Public Goods: One Versus Two Provision Levels," *Journal of Conflict Resolution*, 2015, *59* (7), 1273–1300.

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

Pástor, L'uboš and Pietro Veronesi, "Inequality aversion, populism, and the backlash against globalization," *The Journal of Finance*, 2021, *76* (6), 2857–2906.

Pustejovsky, James E and Elizabeth Tipton, "Meta-Analysis with Robust Variance Estimation: Expanding the Range of Working Models," *Prevention Science*, 2022, *23* (3), 425–438.

Riedl, Arno and Paul Smeets, "Why do investors hold socially responsible mutual funds?," *the Journal of Finance*, 2017, *72* (6), 2505–2550.

Riley, Richard D, "Multivariate Meta-Analysis: the Effect of Ignoring Within-Study Correlation," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2009, *172* (4), 789–811.

_ , KR Abrams, PC Lambert, AJ Sutton, and JR Thompson, "An Evaluation of Bivariate Random-Effects Meta-Analysis for the Joint Synthesis of Two Correlated Outcomes," *Statistics in Medicine*, 2007, *26* (1), 78–97.

Roth, Alvin E and Ido Erev, "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 1995, *8* (1), 164–212.

Sáez, Ignacio, Lusha Zhu, Eric Set, Andrew Kayser, and Ming Hsu, "Dopamine Modulates Egalitarian Behavior in Humans," *Current Biology*, 2015, *25* (7), 912–919.

Sera, Francesco, Benedict Armstrong, Marta Blangiardo, and Antonio C Gasparrini, "An Extended Mixed-Effects Framework for Meta-Analysis," *Statistics in Medicine*, 2019, p. DOI: 10.1002/sim.8362.

Simonsohn, Uri, Leif D Nelson, and Joseph P Simmons, "P-Curve: a Key to the File-Drawer," *Journal of Experimental Psychology: General*, 2014, *143* (2), 534.

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

_ and Hristos Doucouliagos, *Meta-Regression Analysis in Economics and Business*, routledge, 2012.

_ and _ , "Neither Fixed nor Random: Weighted Least Squares Meta-Regression," *Research Synthesis Methods*, 2017, *8* (1), 19–42.

Teyssier, Sabrina, "Inequity and Risk Aversion in Sequential Public Good Games," *Public Choice*, 2012, *151* (1), 91–119.

Tol, Richard SJ, "International Inequity Aversion and the Social Cost of Carbon," *Climate Change Economics*, 2010, *1* (01), 21–32.

Trikalinos, Thomas A, David C Hoaglin, and Christopher H Schmid, "An Empirical Comparison of Univariate and Multivariate Meta-Analyses for Categorical Outcomes," *Statistics in Medicine*, 2014, *33* (9), 1441–1459.

Vogt, Carsten, "Climate Coalition Formation when Players are Heterogeneous and Inequality Averse," *Environmental and Resource Economics*, 2016, *65* (1), 33–59.

Wu, Yan, Hongbo Yu, Bo Shen, Rongjun Yu, Zhiheng Zhou, Guoping Zhang, Yushi Jiang, and Xiaolin Zhou, "Neural Basis of Increased Costly Norm Enforcement under Adversity," *Social Cognitive and Affective Neuroscience*, 2014, *9 (12)*, 1862–1871.

Yamagishi, Toshio and Toko Kiyonari, "The Group as the Container of Generalized Reciprocity," *Social Psychology Quarterly*, 2000, pp. 116–132.

Yang, Yang, Sander Onderstal, and Arthur Schram, "Inequity Aversion Revisited," *Journal of Economic Psychology*, 2016, *54*, 1–16.

Zelmer, Jennifer, "Linear Public Goods Experiments: A Meta-Analysis," *Experimental Economics*, 2003, *6* (3), 299–310.

A Articles Included in Dataset (Chronological Order)

1. **Fehr, Ernst, and Klaus M. Schmidt**, "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 1999, 114(3): 817–868.

2. **Goeree, Jacob, and Charles Holt**, "Asymmetric Inequality Aversion and Noisy Behavior in Alternating-Offer Bargaining Games," *European Economic Review*, 2000, 44(4-6): 1079–1089.

3. **Huck, Steffen, Wieland Müller, and Hans-Theo Normann**, "Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets," *The Economic Journal*, 2001, 111(474): 749–765.

4. **Charness, Gary, and Hernan Haruvy**, "Altruism, Equity, and Reciprocity in a Gift-Exchange Experiment: An Encompassing Approach," *Games and Economic Behavior*, 2002, 40(2): 203–231.

5. **Charness, Gary, and Matthew Rabin**, "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, 2002, 117(3): 817–869.

6. **Ellingsen, Tore, and Magnus Johannesson**, "Promises, Threats and Fairness," *The Economic Journal*, 2004, 114(495): 397–420.

7. **Dannenberg, Astrid, Thomas Riechmann, Bodo Sturm, and Carsten Vogt**, "Inequity Aversion and Individual Behavior in Public Good Games: An Experimental Investigation," Unpublished Manuscript, 2007.

8. **Bellemare, Charles, Sabine Kröger, and Arthur Van Soest**, "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," *Econometrica*, 2008, 76(4): 815–839.

9. **Chen, Yan, and Sherry Xin Li**, "Group Identity and Social Preferences," *American Economic Review*, 2009, 99: 431–457.

10. **Cabrales, Antonio, Raffaele Miniaci, Marco Piovesan, and Giovanni Ponti**, "Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts," *American Economic Review*, 2010, 100: 2261–2278.

11. **Dannenberg, Astrid, Bodo Sturm, and Carsten Vogt**, "Do Equity Preferences Matter for Climate Negotiators? An Experimental Investigation," *Environmental and Resource Economics*, 2010, 47(1): 91–109.

12. **Daruvala, Dinky**, "Would the Right Social Preference Model Please Stand Up!," *Journal of Economic Behavior & Organization*, 2010, 73(2): 199–208.

13. **Lau, Sau-Him Paul, and Felix Leung**, "Estimating a Parsimonious Model of Inequality Aversion in Stackelberg Duopoly Experiments," *Oxford Bulletin of Economics and Statistics*, 2010, 72(5): 6690–686.

14. **Messer Kent D., Gregory L. Poe, Daniel Rondeau, William D. Schulze, and Christian A. Vossler**, "Social Preferences and Voting: An Exploration using a Novel Preference Revealing Mechanism," *Journal of Public Economics*, 2010, 94 (3-4): 308–317.

15. **Bellemare, Charles, Sabine Kröger, and Arthur Van Soest**, "Preferences, Intentions, and Expectation Violations: A Large-Scale Experiment with a Representative Subject Pool," *Journal of Economic Behavior & Organization*, 2011, 78:(3): 349–365.

16. **Blanco, Mariana, Dirk Engelmann, and Hans-Theo Normann**, "A Within-Subject Analysis of Other-Regarding Preferences," *Games and Economic Behavior*, 2011, 72(2): 321–338.

17. **Wright, Nicholas D., Mkael Symmonds, Stephen M. Fleming, and Raymond J. Dolan**, "Neural Segregation of Objective and Contextual Aspects of Fairness," *Journal of Neuroscience*, 2011, 31 (14): 5244-5252.

18. **Morishima, Yosuke, Daniel Schunk, Adrian Bruhin, Christian C.Ruff, Ernst Fehr**, "Linking Brain Structure and Activation in Temporoparietal Junction to Explain the Neurobiology of Human Altruism," *Neuron*, 2012, 75(1): 73–79.

19. **Aksoy, Ozan, and Jeroen Weesie**, "Hierarchical Bayesian Analysis of Biased Beliefs and Distributional Other-Regarding Preferences," *Games*, 2013, 4(1): 66–88.

20. **Kleine, Fabian, Manfred Königstein, and Balázs Rozsnyói**, "Voluntary Leadership in an Experimental Trust Game," *Journal of Economic Behavior & Organization*, 2014, 108: 442–452.

21. **Wu, Yan, Hongbo Yu, Bo Shen, Rongjun Yu, Zhiheng Zhou, Guoping Zhang, Yushi Jiang, and Xiaolin Zhou**, "Neural Basis of Increased Costly Norm Enforcement under Adversity," *Social Cognitive and Affective Neuroscience*, 2014, 9 (12): 1862-1871.

22. **Beranek, Benjamin, Robin Cubitt, and Simon Gächter**, "Stated and Revealed Inequality Aversion in Three Subject Pools" *Journal of the Economic Science Association*, 2015, 1: 43–58.

23. **Corgnet, Brice, Antonio M. Espín, and Roberto Hernán-González**, "The Cognitive Basis of Social Behavior: Cognitive Reflection Overrides Antisocial but not Always Prosocial Motives," *Frontiers in Behavioral Neuroscience*, 2015.

24. **Ponti, Giovanni, Ismael Rodriguez-Lara**, "Social Preferences and Cognitive Reflection: Evidence from a Dictator Game Experiment," *Frontiers in Behavioral Neuroscience*, 2015.

25. **Saez, Ignacio, Lusha Zhu, Eric Set, Andrew Kayser, and Ming Hsu**, "Dopamine Modulates Egalitarian Behavior In Humans," *Current Biology*, 2015, 30;25 (7): 912-919.

26. **He, Haoran, Keyu Wu**, “Choice Set, Relative Income, and Inequity Aversion: An Experimental Investigation,” *Journal of Economic Psychology*, 2016, 54: 177–193.

27. **Cueva, Carlos, Iñigo Iturbe-Ormaetxe, Esther Mata-Pérez, Giovanni Ponti, Marcello Sartarelli, Haihan Yu, and Vita Zhukova**, “Cognitive (Ir)Reflection: New Experimental Vvidence,” *Journal of Behavioral and Experimental Economics*, 2016, 64: 81–93.

28. **De Melo, Celso, Stacy Marsella, and Jonathan Gratch**, “People Do Not Feel Guilty About Exploiting Machines,” *ACM Transactions on Computer-Human Interaction*, 2016, 23(2): 1–17.

29. **Yang, Yang, Sander Onderstal, and Arthur Schram**, “Inequity Aversion Revisited,” *Journal of Economic Psychology*, 2016, 54: 1–16.

30. **Boun My, Kene, Nicolas Lampach, Mathieu Lefebvre, and Jacopo Magnani**, “Effects of Gain-Loss Frames on Advantageous Inequality Aversion,” *Journal of the Economic Science Association*, 2018, 4(2): 99–109.

31. **Hu, Yang, Lisheng He, Lei Zhang, Thorben Wölk, Jean-Claude Dreher, and Bernd Weber**, “Spreading Inequality: Neural Computations Underlying Paying-It-Forward Reciprocity,” *Social Cognitive and Affective Neuroscience*, 2018, 13 (6): 578-589.

32. **Tasch, Weiwei, and Daniel Houser**, “Social Preferences and Social Curiosity,” Unpublished Manuscript, 2018.

33. **Gao, Xiaoxue, Hongbo Yu, Ignacio Sáez, Philip R Blue, Lusha Zhu, Ming Hsu, and Xiaolin Zhou**, “Distinguishing Neural Correlates of Context-dependent Advantageous- and Disadvantageous-inequity Aversion,” *Proceedings of the National Academy of Sciences*, 2018, 115 (33): E7680-E7689.

34. **Bruhin, Adrian, Ernst Fehr, Daniel Schunk**, “The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences,” *Journal of the European Economic Association*, 2019, 17(4): 1025–1069.

35. **Müller, Stephan, and Holger A. Rau**, “Decisions under Uncertainty in Social Contexts,” *Games and Economic Behavior*, 2019, 116: 73–95.

36. **Kerschbamer, Rudolf, and Daniel Müller**, “Social Preferences, Political Opinions and Charitable Giving: An Online Experiment on a Large Heterogeneous Sample,” *Journal of Public Economics*, 2020, 182.

37. **Hedegaard, Morten, Rudolf Kerschbamer, Daniel Müller, and Jean-Robert Tyran**, “Distributional Preferences Explain Individual Behavior across Games and Time,” *Games and Economic Behavior*, 2021, 128: 231–255.

38. **Sabater-Grande, Gerardo, Aurora García-Gallego, Nikolaos Georgantzís, and Noemi Herranz-Zarzoso**, "The Effects of Personality, Risk and Other-Regarding Attitudes on Trust and Reciprocity," *Journal of Behavioral and Experimental Economics*, 2022, 96.

39. **Diaz, Lina, Daniel Houser, John Ifcher, and Homa Zarghamee**, "Estimating Social Preferences Using Stated Satisfaction: Novel Support for Inequity Aversion," *European Economic Review*, 2023, 155: 104436.

40. **Alger, Ingela, and Boris van Leeuwen**, "Estimating Social Preferences and Kantian Morality in Strategic Interactions," *Journal of Political Economy Microeconomics*, 2024, 2: 665–706.

41. **Carpenter, Jeffrey, and Andrea Robbett**, "Measuring Socially Appropriate Social Preferences," *Games and Economic Behavior*, 2024, 147: 517–532.

B Variables Coded in the Dataset

Table 14: List of Coded Variables in the Dataset

Variable	Description
study_id	ID for the 43 studies in the analysis (from 1 to 43)
paper_title	Title of the paper
authors	Authors' first and last names
paper_code	First author's last name + et al. + year
is_published	= 1 if the paper is published
year_published	Year published or last revisited if working paper
journal	Journal
paper_length	Length of the paper (appendix excluded)
affiliations	Affiliations of the authors
is_lab	= 1 if laboratory experiment
is_online	= 1 if online experiment
is_classroom	= 1 if classroom experiment
loc_exp_country	Country location of the experiment
loc_exp_contintent	Continent location of the experiment
is_uni	= 1 if university students population
is_adults	= 1 if adults population (not general or in university)
is_general	= 1 if general population
reward_money	= 1 if monetary reward
strategic_alpha	= 1 if α elicited in a strategic game
strategic_beta	= 1 if β elicited in a strategic game
games_alpha	Games used to elicit α
games_beta	Games used to elicit β
game1-game4	All games played in the experiment
utility_function	Utility function specification used
econometric_strategy	Econometric strategy
estimation_method	Estimation method used
alpha	Disadvantageous inequality coefficient (α)
alpha_se	SE of α
alpha_sd	SD of α
beta	Advantageous inequality coefficient (β)
beta_se	SE of β
beta_sd	SD ofβ
type_se	Type of SE (reported, from SD, from reg)
type_sd	Type of SD (reported, computed)

n	Sample size
is_aggregate	= 1 if aggregate estimates
is_individual	= 1 if individual-level estimates
is_mean	= 1 if individual-level mean
is_median	= 1 if individual-level median
is_finite_mix	= 1 if finite-mixture estimates
p1-p4	mixture probabilities if finite-mixture
p1_se-p4_se	SEs of $p_1 - p_4$ if finite-mixture
alpha1-alpha4	Alpha coefficients if finite-mixture
alpha1_se-alpha4_se	SEs of $\alpha_1 - \alpha_4$ if finite-mixture
beta1-beta4	Beta coefficients if finite-mixture
beta1_se-beta4_se	SEs of $\beta_1 - \beta_4$ if finite-mixture
t-stat	t-statistics of the estimate
is_other_param	= 1 if other parameters are estimated
other_param	Names of other parameters
other_info	Other information on the paper

C Experimental Tasks Used To Elicit Parameters

Table 15: Experimental Tasks and Classification as Strategic

Experimental Tasks Used To Elicit Parameters	Strategic Environment
Disadvantageous Inequality Coefficient (α)	
Alternating-offer bargaining game	Yes
Dictator game	No
Equality equivalence test	No
Gift exchange game	Yes
Mini dictator game	No
Mini dictator game with equality-efficiency trade-off	No
Production game	Yes
Public good game with voting mechanism	Yes
Response game	Yes
Reciprocity game	Yes
Sequential prisoner dilemma	Yes
Sequential public good game	Yes
Simultaneous production game	Yes
Stackelberg duopoly game	Yes
Trust game	Yes
Ultimatum game	Yes
Advantageous Inequality Coefficient (β)	
Alternating-offer bargaining game	Yes
Dictator game	No
Equality equivalence test	No
Gift exchange game	Yes
Mini dictator game	No
Mini dictator game with equality-efficiency trade-off	No
Production game	Yes
Public good game with voting mechanism	Yes
Response game	Yes
Reciprocity game	Yes
Sequential prisoner dilemma	Yes
Sequential public good game	Yes
Simultaneous production game	Yes
Stackelberg duopoly game	Yes
Trust game	Yes
Ultimatum game	Yes

D Meta-Analysis with Bayesian Hierarchical Model

We now explain the modelling framework of the Bayesian hierarchical model. We will use in the examples the variable α, but the same applies also to β. Consider the dataset $(\alpha_j, se_j^2)_{j=1}^k$, where k is the total number of estimates and α_j the jth observation of the disadvantageous inequality parameter, with its associated standard error se_j. We then assume that the reported estimate α_j is distributed normally around the parameter $\bar{\alpha}_j$:

$$\alpha_j|\bar{\alpha}_j, se_j \sim \mathcal{N}(\bar{\alpha}_j, se_j^2)$$

The variability around $\bar{\alpha}_j$ is due to the sampling variation captured by the standard errors se_j. As in a frequentist random-effects model, we can assume that the sampling variation is not the only source of variability for the estimates, since there could be heterogeneity across measurements due to different settings like subject population, games played etc. This can be modeled by assuming that each $\bar{\alpha}_j$ is normally distributed, adding a second layer to the hierarchy:

$$\bar{\alpha}_j|\alpha_0, \tau \sim \mathcal{N}(\alpha_0, \tau^2)$$

where α_0 is the overall mean of the disadvantageous inequality parameters $\bar{\alpha}_j$, and τ^2 represents the genuine variability across studies. Combining the two expressions we get:

$$\alpha_j|\alpha_0, \tau, se_j \sim \mathcal{N}(\alpha_0, \tau^2 + se_j^2)$$

In Bayesian hierarchical models, each observation, $\bar{\alpha}_j$, is pooled towards the overall mean with strength depending on the precision of the estimate and on how far the estimate is from the α_0. The pooling equation can be written as follows:

$$\bar{\alpha}_j = (1 - \omega_j)\alpha_j + \omega_j \alpha_0$$

where ω_j is the "pooling factor" (Gelman and Pardoe, 2006), defined as:

$$\omega_j = \frac{se_j^2}{\tau^2 + se_j^2}$$

All others things considered, the more an estimate is imprecise, captured by se_j, the more it will be pooled towards the overall mean. The same effect also happens when τ^2 is low, meaning that if there is low heterogeneity across studies, more weight will be given to α_0.

The model above does not take into account the possibility of statistically dependent estimates, for example for estimates that come from the same study. One way to address this problem is to introduce a paper level, that captures the mean of the parameter in a single study, $\bar{\alpha}_p$, and the variability within study σ_p^2. These models for α and β resemble the three-level frequentist approach discussed in details in the main body of the paper. Overall, the models can be written as follow:

$$\begin{aligned}
\alpha_{pj}|\bar{\alpha}_{pj}, se_{pj} &\sim \mathcal{N}(\bar{\alpha}_{pj}, se_{pj}^2) \\
\bar{\alpha}_{pj}|\bar{\alpha}_p, \sigma_p &\sim \mathcal{N}(\bar{\alpha}_p, \sigma_p^2) \\
\bar{\alpha}_p|\alpha_0, \tau_s &\sim \mathcal{N}(\alpha_0, \tau_s^2) \\
\alpha_0 &\sim \mathcal{N}(0.25, 1) \\
\tau &\sim \text{half } \mathcal{N}(0, 1) \\
\sigma_p &\sim \text{half } \mathcal{N}(0, 1)
\end{aligned}
\qquad
\begin{aligned}
\beta_{pj}|\bar{\beta}_{pj}, se_{pj} &\sim \mathcal{N}(\bar{\beta}_{pj}, se_{pj}^2) \\
\bar{\beta}_{pj}|\bar{\beta}_p, \sigma_p &\sim \mathcal{N}(\bar{\beta}_p, \sigma_p^2) \\
\bar{\beta}_p|\beta_0, \tau &\sim \mathcal{N}(\beta_0, \tau^2) \\
\beta_0 &\sim \mathcal{N}(0.25, 1) \\
\tau &\sim \text{half } \mathcal{N}(0, 1) \\
\sigma_p &\sim \text{half } \mathcal{N}(0, 1)
\end{aligned}$$

We now summarize and estimate the models expressed above. We estimate the models in Stan (Carpenter et al., 2017) using the Hamiltonian Monte Carlo simulations and launch it from R (https:// www.r-project.org/) using RStan (Stan Development Team, 2021).

The priors for the population parameters are mildly regularising, meaning that they are informative but are chosen in such a way to have a weak effect in the procedure. Looking, for example, at the prior for α_0 and by using the three sigma-rule of thumb, what the prior is saying is that our initial opinion for the true value of α_0 is that the parameter lies between -1.75 and 2.25 with 95% probability. The procedure is not sensitive to the priors we use as long as they are weakly informative.

The Bayesian procedure returns a mean disadvantageous inequality coefficient of 0.530, with a 95% probability that the true value falls in the interval [0.310, 0.751]. This is in line with what we found in the frequentist analysis, with an estimate for α of 0.533 and a confidence interval of [0.311, 0.755].

Table 16: Summary of the Bayesian Hierarchical Model Estimate for α with paper level

Parameter	Rhat	ESS	Mean	SD	2.5%	25%	50%	75%	97.5%
α_0	1.000	23121	0.530	0.112	0.310	0.454	0.530	0.604	0.751
$\hat{\tau}^2$	1.000	16680	0.522	0.128	0.327	0.432	0.504	0.594	0.830

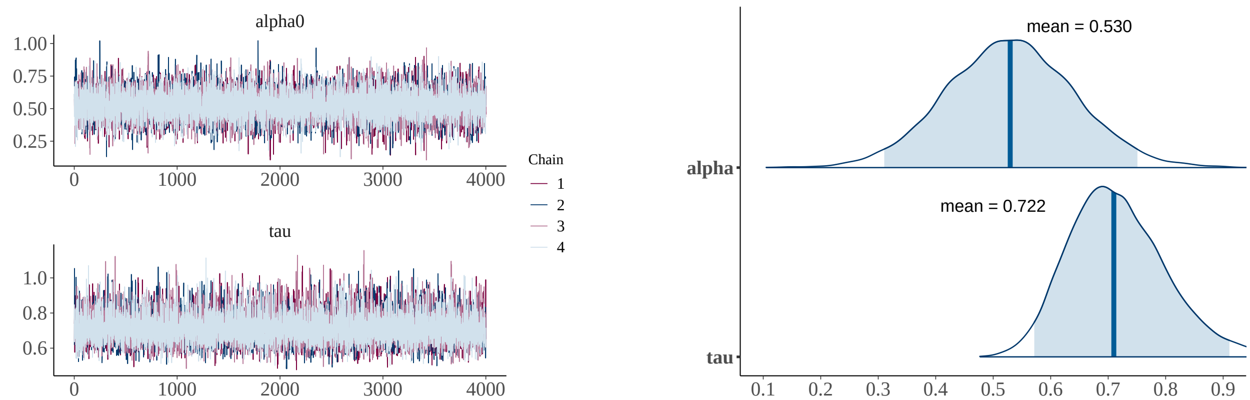


Figure 11: The first figure shows the 1,500 draws in the four Markov Chain after the warmup, showing good convergence of the procedure. The second figure shows the posterior distributions of α_0 and τ. Shaded blue areas correspond to 95% credible intervals.

Now discussing β, the Bayesian procedure returns a mean advantageous inequality coefficient of 0.326, with a 95% probability that the true value falls in the interval [0.253, 0.398]. Once again, this is in line with what we found in the frequentist analysis, with an estimate for β of 0.326 and a confidence interval of [0.254, 0.398].

Table 17: Summary of the Bayesian Hierarchical Model Estimate for β with paper level

Parameter	Rhat	ESS	Mean	SD	2.5%	25%	50%	75%	97.5%
β_0	1.000	16185	0.326	0.037	0.253	0.301	0.326	0.350	0.398
$\hat{\tau}^2$	1.000	11795	0.043	0.013	0.024	0.034	0.042	0.050	0.074

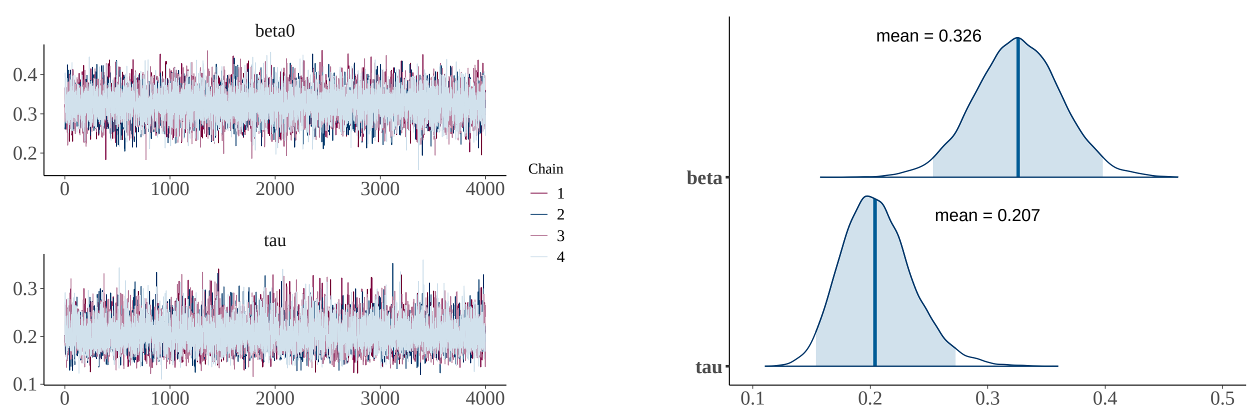


Figure 12: The first figure shows the 1,500 draws in the four Markov Chain after the warmup, showing good convergence of the procedure. The second figure shows the posterior distributions of α_0 and τ. Shaded blue areas correspond to 95% credible intervals.

We provide here a table with a sensitivity analysis on the priors chosen for the bayesian models by looking at the average of the parameters and their 95% credible intervals.

Table 18: Sensitivity Analysis on Priors for the three-Level Model

Prior	Disadvantageous Inequality (α_0)		Advantageous Inequality (β_0)	
$sd = 2, \phi_p = 0.25$	0.531	[0.302,0.758]	0.330	[0.267,0.393]
$sd = 0.5, \phi_p = 0.25$	0.519	[0.311,0.729]	0.330	[0.271,0.391]
$sd = 1, \phi_p = 0$	0.526	[0.304,0.747]	0.330	[0.267,0.392]
	Average	95% Credible	Average	95% Credible

Notes: sd is the standard deviation used for all priors. ϕ_p is the mean of the normal prior on the parameter, for both α and β.

The previous two modelling frameworks assume normality in all the levels. While this assumption seems reasonable given the empirical distribution of estimates that we get for β, the same might not be true for the other parameter α. In fact, Figure 1 shows a distribution that is right-skewed, and the normality assumption might cause the construction of imprecise credible intervals for the estimate due to the inherent symmetry of the distribution. Given the parameters' values we found for α and τ^2 assuming a normal population level, we would be slightly underestimating values of α around zero, and slightly overestimating the left-tail of the distribution. To solve this problem, and as an additional sensitivity analysis, we estimate the following model for α:

$$\begin{aligned}
\alpha_{pj}|\bar{\alpha}_{pj}, se_{pj} &\sim \mathcal{N}(\bar{\alpha}_{pj}, se_{pj}^2) \\
\bar{\alpha}_{pj}|\bar{\alpha}_p, \sigma_p &\sim \mathcal{N}(\bar{\alpha}_p, \sigma_p^2) \\
\bar{\alpha}_p|\xi, \omega, \theta &\sim Skew - \mathcal{N}(\xi, \omega, \theta) \\
\xi &\sim \mathcal{N}(0.25, 1) \\
\omega &\sim \text{half } \mathcal{N}(0, 1) \\
\theta &\sim \mathcal{N}(5, 5) \\
\sigma_p &\sim \text{half } \mathcal{N}(0, 1)
\end{aligned}$$

where we assume that the population level is distributed as a Skew-Normal distribution. The density of a Skew-Normal distribution is:

$$f(x) = \frac{2}{\omega\sqrt{2\pi}} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\theta(\frac{x-\xi}{\omega})} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

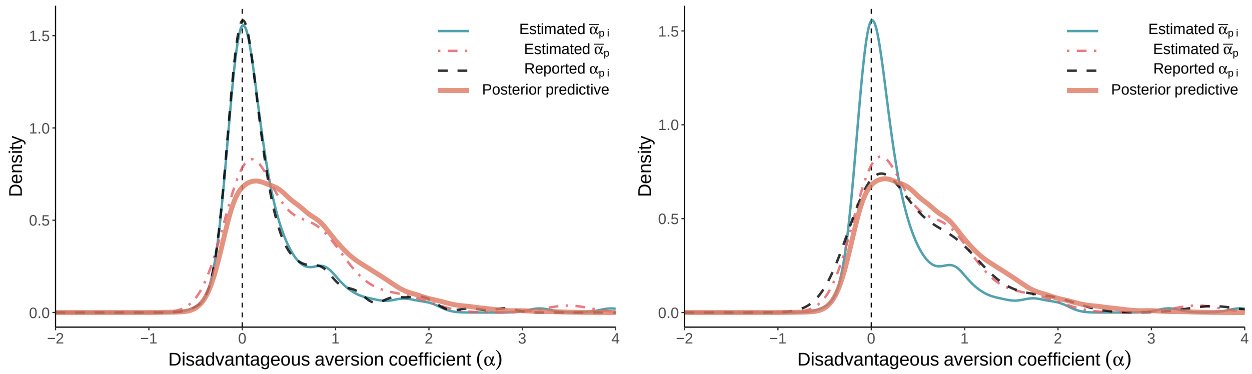


Figure 13: Distribution of empirical, estimated and predictive distribution of the Disadvantageous Inequality Coefficient (α). The left Figure plots all the reported estimates for α in our dataset. The right Figure plots only one estimate per paper. In case the paper reports multiple estimate, we compute and consider only the mean.

with ξ being a location parameter, ω being a scale parameter, and θ being a shape parameter. If $\theta = 0$, we go back to a normal distribution, if $\theta > 0$ we get a right-skewed distribution, while if $\theta < 0$ we get a left-skewed distribution. We can recover the mean and variance of the distribution, what we are interested in, from the parameters using the following formulas:

$$\text{Mean} = \alpha_0 = \xi + \omega\delta\sqrt{\frac{2}{\pi}} \qquad \text{Variance} = \tau^2 = \omega^2\left(1 - \frac{2\delta^2}{\pi}\right)$$

with $\delta = \frac{\theta}{\sqrt{1+\theta^2}}$. Moreover, by relaxing the normality assumption at the population level, we can compute additional distributional measures of interest, such as the median. As we can see from Table 19, the mean value for α in the population is 0.609, with a 95% credible interval of [0.433, 0.818], thus on average we see inequality aversion in the disadvantageous realm also in this specification. This holds true also for the median, with a value of 0.488 and 95% credible interval of [0.312, 0.697]. By looking at Figure 13 we see that the predictive distribution fits well the right-skewed empirical distribution for α.

We now estimate the same model in the subsamples of strategic and non-strategic environments to check whether we obtain the same conclusion of inequality aversion.

Referring to Table 19, we once again observe a positive value for both the mean and the median, with 95% credible intervals in strategic environments of [0.522, 1.161] and [0.380, 1.019] respectively. Similar conclusions hold when focusing on non-strategic environments, with credible intervals of [0.145, 0.519] for the mean and [0.078, 0.452] for the median. Overall, these results suggest that the skewness of the empirical distribution and the normality assumption were not driving the results towards behindness aversion.

Table 19: Estimation for Disadvantageous Inequality (α) – Bayesian Hierarchical Model

Pop. Distribution	**Sample**	**Measure**	**Parameter**	**95% Cred. Interval**
Skew-Normal	All	Mean	$\alpha = 0.609$	[0.433, 0.818]
Skew-Normal	NS	Mean	$\alpha = 0.305$	[0.145, 0.519]
Skew-Normal	S	Mean	$\alpha = 0.810$	[0.522, 1.161]
Skew-Normal	All	Median	$\alpha = 0.488$	[0.312, 0.697]
Skew-Normal	NS	Median	$\alpha = 0.238$	[0.078, 0.452]
Skew-Normal	S	Median	$\alpha = 0.668$	[0.380, 1.019]

Notes: "NS" denotes the non-strategic subsample, while "S" represents the strategic subsample. Since the median of a skew-normal distribution does not have a closed-form expression, we approximate its 95% credible interval using the credible interval for the mean. Specifically, we compute the distance between the mean point estimate and its lower bound and subtract this from the median point estimate to obtain an approximate lower bound for the median. Analogously, we compute the distance between the mean point estimate and its upper bound and add this to the median point estimate to approximate the upper bound for the median.

E Multivariate Meta-Analysis

When conducting meta-analyses encompassing studies that report multiple effect sizes, various methodological approaches are available. The first one is to consider each effect size independent of the others and conduct univariate analysis, one for each effect size. Univariate meta-analysis are simple to implement and interpret, but this approach completely disregards possible within-study and between-study outcome correlations that can have a potentially relevant effect on the estimates and their SEs.

The alternative approach is to implement a multivariate meta-analysis by explicitly modelling outcomes correlations'. While multivariate models are theoretically the first-best, since they can always nest univariate models, they are more difficult and time-consuming to estimate. Furthermore, empirical investigations conducted by (Trikalinos et al., 2014; Berkey et al., 1998; Ishak et al., 2008) find little to no effect on the parameter estimates between univariate and multivariate meta-analysis, thus supporting the idea of simply using the easier univariate model. Other studies (Riley et al., 2007; Kirkham et al., 2012) find instead a difference between univariate and multivariate estimates, and they argue that a multivariate approach is the correct procedure when dealing with multiple effect sizes in the same study.

Another problem in conducting a multivariate meta-analysis is the need to not only have a measure of the effect sizes and their SEs, but also of their correlation (or covariance), and this information is often not reported. Ishak et al. (2008) suggest that the correlation can be ignored without too much risk of introducing a bias in the analysis, but Riley (2009) finds that this was not true in the papers he analysed.

The specification for the multivariate random-effects model applied in our dataset of inequality sensitivity estimates is the following:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} \mu_j^{\alpha} \\ \mu_j^{\beta} \end{pmatrix}, R_j \right\}, \qquad R_j = \begin{bmatrix} SE_{aj}^2 & SE_{aj}SE_{bj}\rho \\ SE_{aj}SE_{bj}\rho & SE_{bj}^2 \end{bmatrix}$$

$$\begin{pmatrix} \mu_j^{\alpha} \\ \mu_j^{\beta} \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix}, D \right\}, \qquad D = \begin{bmatrix} D_a^2 & D_a D_b \rho_D \\ D_a D_b \rho_D & D_b^2 \end{bmatrix}$$

Where similarly to the univariate model, we assume that the observed parameters (α_j, β_j) are distributed around the true effect sizes $(\mu_j^{\alpha}, \mu_j^{\beta})$, with known variance-covariance matrix R_j. The diagonal elements are the variance for α and β which are known, while ρ is the correlation among the estimates in our dataset. As we do not possess information on the correlation (or covariance), we will do a sensitivity analysis by assuming different values of

ρ, constant across estimates. The true effect sizes are then distributed as a bivariate normal with means (α_0, β_0) and variance-covariance matrix D.

To handle statistically dependent estimates we can add another level to the hierarchy to capture both within-study and between-study heterogeneity, thus getting a multivariate and three-level specification:

$$\begin{pmatrix} \alpha_{ij} \\ \beta_{ij} \end{pmatrix} \sim \mathcal{N}\left\{\begin{pmatrix} \mu^{\alpha}_{ij} \\ \mu^{\beta}_{ij} \end{pmatrix}, R_{ij}\right\}, \qquad R_{ij} = \begin{bmatrix} SE^2_{aij} & SE_{aij}SE_{bij}\rho \\ SE_{aij}SE_{bij}\rho & SE^2_{bij} \end{bmatrix}$$

$$\begin{pmatrix} \mu^{\alpha}_{ij} \\ \mu^{\beta}_{ij} \end{pmatrix} \sim \mathcal{N}\left\{\begin{pmatrix} \theta^{\alpha}_{i} \\ \theta^{\beta}_{i} \end{pmatrix}, C\right\}, \qquad C = \begin{bmatrix} C^2_{a} & C_a C_b \rho_C \\ C_a C_b \rho_C & C^2_{b} \end{bmatrix}$$

$$\begin{pmatrix} \theta^{\alpha}_{i} \\ \theta^{\beta}_{i} \end{pmatrix} \sim \mathcal{N}\left\{\begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix}, D\right\}, \qquad D = \begin{bmatrix} D^2_{a} & D_a D_b \rho_D \\ D_a D_b \rho_D & D^2_{b} \end{bmatrix}$$

Where the observed parameters $(\alpha_{ij}, \beta_{ij})$ are distributed around the true effect sizes $(\mu^{\alpha}_{ij}, \mu^{\beta}_{ij})$, the true effect sizes around paper-level means $(\theta^{\alpha}_{i}, \theta^{\beta}_{i})$ and the latter around the population means (α_0, β_0). In this multivariate three-level model we are estimating in addition to the variance of the within and between study errors for α and β, also their correlation/covariance.

We report in Table 20 the results of the multivariate three-level model for different values of the correlation coefficient ρ, estimated in R using the package "mixmeta" Sera et al. (2019). As we can see, the mean values for the parameters of interest α and β are extremely similar to the ones obtained in the univariate model, where $\alpha = 0.533$ and $\beta = 0.326$.

It is interesting now to look at the correlation coefficients at the population level and paper level, which are one of the main gains of running a multivariate model. The correlation at the population level ρ_D is close to zero for any ρ that we assumed at the higher level, with values ranging between 0.08 and 0.05. Looking now at the paper level correlation ρ_C, we see more heterogeneity depending on the underlying assumption on ρ. If we focus on plausible values of ρ based on the information in our dataset and in the literature, namely values between [-0.2, 0.2], we see a negative correlation ρ_C between α and β. In the Fehr and Schmidt (1999) model, a positive correlation implies that agents who dislike having more, dislike also having less. With a negative correlation, we see instead that strong advantageous inequality β is associated with lower values of α. This means that people with higher values of the guilt parameter β, are also the ones less envious.

Table 20: Multivariate Three-level Model for α and β

ρ	α	β	C_a^2	C_b^2	ρ_C	D_a^2	D_b^2	ρ_D
-0.8	0.527	0.327	0.043	0.022	0.096	0.497	0.04	0.083
	(0.111)	(0.036)						
-0.6	0.531	0.328	0.041	0.022	-0.002	0.496	0.04	0.075
	(0.110)	(0.036)						
-0.4	0.533	0.328	0.041	0.022	-0.088	0.495	0.04	0.069
	(0.110)	(0.036)						
-0.2	0.535	0.328	0.041	0.022	-0.173	0.494	0.04	0.063
	(0.110)	(0.036)						
0.0	0.537	0.327	0.040	0.023	-0.259	0.494	0.04	0.059
	(0.110)	(0.036)						
0.2	0.540	0.327	0.041	0.023	-0.349	0.494	0.04	0.054
	(0.110)	(0.036)						
0.4	0.543	0.326	0.042	0.024	-0.444	0.493	0.04	0.050
	(0.110)	(0.036)						
0.6	0.546	0.325	0.043	0.026	-0.546	0.493	0.04	0.047
	(0.110)	(0.036)						
0.8	0.551	0.323	0.047	0.030	-0.655	0.492	0.04	0.045
	(0.110)	(0.037)						

F Robustness Checks

F.1 Normalization of Parameter Space

The range of the parameters α and β is asymmetric. Specifically, $\alpha \in [-1, \infty)$ and $\beta \in (-\infty, 1]$, with the boundaries at -1 and 1 preventing situations where an individual would be willing to "burn money" to increase the payoff gap when behind or reduce it when ahead. This asymmetry in the parameter range raises concerns about the potential for a mechanical bias in our weighted mean estimates, particularly for α, which exhibits a right-skewed distribution with a significant number of estimates greater than 1. In this Appendix, we revisit the interpretation of the α parameter and present two robustness checks to address this asymmetry in the parameter range. Together, these robustness checks suggest that our findings are not an artifact of the asymmetry in the range of α.

Behavioral Interpretation of α

When i faces disadvantageous inequality (i.e., when $x_i < x_j$), the FS model reduces to:

$$U_i(x_i, x_j) = x_i - \alpha(x_j - x_i) = (1 + \alpha)x_i - \alpha x_j,$$

where x_i is the decision-maker's payoff, x_j is the other player's payoff, and α captures the decision-maker's sensitivity to disadvantageous inequality.

If α is in the $[-1, 0)$ range, the individual is altruistic, placing positive weight on the other's payoff. This implies a willingness to sacrifice part of their own payoff to increase the other's. Imposing a lower bound equal to -1 is meant to exclude cases where one's own payoff becomes a "bad," though the model could, in principle, accommodate such behavior. Thus, the range $[-1, 0)$ adequately captures altruism and the relevant trade-offs. Conversely, when $\alpha > 0$, the individual is envious, placing negative weight on the other's payoff. This implies a willingness to sacrifice part of their own payoff to reduce the other's. FS do not impose an upper bound here, as extreme envy can capture a strong negative reaction to others' earnings without violating any axiom or desirable feature of preferences (for example, one's own earnings being a "good" rather than a "bad" regardless of the other's earnings). Once again, the model can capture such trade-offs in the range $(0, +\infty)$.

In short, the asymmetry in the parameter space does not constrain the range of behaviors the model can capture: we are still able to observe the full range of altruism and envy within the FS framework. The restriction on α is simply meant to exclude cases where the individual treats their own earnings as a bad.

Table 21: Estimation for Disadvantageous Inequality (α') – Normalization

Sample	Parameter	SE	P-value	Nr Est	Nr Studies
All	$\alpha' = 0.530$	0.110	< 0.001	149	43
NS	$\alpha' = 0.253$	0.096	0.017	65	18
S	$\alpha' = 0.730$	0.169	< 0.001	84	25

Notes: In this Table negative estimates are normalized such that $\alpha' = \alpha/(1+\alpha)$, while positive estimates remain the same $\alpha' = \alpha$. "NS" denotes the non-strategic subsample, while "S" represents the strategic subsample.

Normalization of Parameter Space, First Transformation

To address the asymmetry at its root, we applied the transformation:

$$\alpha' = \frac{\alpha}{1+\alpha},$$

to the negative values of α. This maps the original range from $[-1, \infty)$ to $(-\infty, \infty)$. Table 21 presents the results using these transformed values of α. The meta-analytic synthesis on the transformed values is consistent with our original findings:

- Full sample: 0.530 (SE: 0.110, $p < 0.001$)
- Non-strategic subsample: 0.253 (SE: 0.096, $p = 0.017$)
- Strategic subsample: 0.730 (SE: 0.169, $p < 0.001$)

The modest effect of the transformation is largely due to the limited presence of large negative α estimates (e.g., the smallest value, -0.455, maps to an α' equal to -0.835). Importantly, this finding is not due to any bounds that the FS model imposes on the untransformed estimates: values of α smaller than -0.455 are feasible within the model, and we would observe such estimates if participants' choices warranted them.

Table 22: Estimation for Willingness to Pay (WTP) in Disadvantageous Situations

Sample	**Parameter**	**SE**	**P-value**	**Nr Est**	**Nr Studies**
All	$WTP = 0.213$	0.038	< 0.001	149	43
NS	$WTP = 0.120$	0.050	0.028	65	18
S	$WTP = 0.283$	0.051	< 0.001	84	25

Notes: In this Table all estimates are normalized such that $\alpha' = \alpha/(1+\alpha)$, which is the WTP to increase/decrease the other's payoff by 1\$. "NS" denotes the non-strategic subsample, while "S" represents the strategic subsample.

Normalization of Parameter Space, Second Transformation

To further address range asymmetry, we reparametrized all values of α using the same transformation as above:

$$\alpha' = \frac{\alpha}{1+\alpha}.$$

This changes the interpretation of the parameter in an interesting way: α' represents the willingness to pay (WTP) to increase (if $\alpha < 0$) or decrease (if $\alpha > 0$) the other participant's payoff by \$1. The formula comes from taking the total derivative of the utility function and setting it equal to zero:

$$U_i(x_i, x_j) = (1+\alpha)x_i - \alpha x_j,$$

$$\frac{dU(x_i, x_j)}{dx} = 0 \quad \Rightarrow \quad (1+\alpha)\, dx_i - \alpha\, dx_j = 0 \quad \Rightarrow \quad dx_i = \frac{\alpha}{1+\alpha}\, dx_j.$$

Setting $dx_j = 1$, we obtain the transformation used.

As a result, the estimate from the meta-analysis summarizes WTPs rather than sensitivities to disadvantageous inequality. The results, reported in Table 22, show that the WTP is positive and significant in all three samples, implying that individuals are willing to pay a strictly positive amount to reduce others' earnings—which is the signature feature of envious distributional preferences and $\alpha > 0$. While this second transformation offers a clear interpretation in terms of willingness to pay (WTP), it also serves as a more stringent test of the sign of the parameter compared to the first transformation. Unlike the first transformation, which merely stretches negative values of α, this re-parameterization also compresses positive values of α.

F.2 Alternative Correlation Assumption: Sample Level

In our main analysis, we assumed that all estimates within the same paper were correlated, due to shared researcher-specific or study-specific characteristics. We perform here a robustness check for α where we classify estimates as independent when drawn from distinct samples even if they are from the same study (e.g., due to different treatments in a between-subjects design, or estimates reported for separate subgroups). Estimates from the same participants (e.g., those generated by varying econometric specifications) are still treated as correlated.

Using this classification, our dataset expands from 43 independent studies to 65 independent samples. Our meta-analysis still yields results in line with those from the main analysis (see Table 23 below):

- Full sample: 0.451 (SE: 0.082, $p < 0.001$)
- Non-strategic subsample: 0.204 (SE: 0.060, $p = 0.002$)
- Strategic subsample: 0.740 (SE: 0.153, $p < 0.001$)

Our conclusions on α are, thus, robust to an alternative correlation structure.

Table 23: Estimation for Disadvantageous Inequality (α) – Sample Level Correlation

Sample	**Parameter**	**SE**	**P-value**	**Nr Est**	**Nr Samples**
All	$\alpha = 0.451$	0.082	< 0.001	149	65
NS	$\alpha = 0.204$	0.06	0.002	65	36
S	$\alpha = 0.740$	0.153	< 0.001	84	29

Notes: In this Table estimates are considered as independent when they derive from distinct samples even if they are from the same study. "NS" denotes the non-strategic subsample, while "S" represents the strategic subsample.

F.3 Correlated Hierarchical Effects (CHE) Model

The three-level random-effects model assumes that, conditional on being in the same study, the parameters are independent. In equation (6), this implies that $Cov(\epsilon_{ij}, \epsilon_{ih}) = 0$ for every estimate $j \neq h$ in study i. The Correlated Hierarchical Effects (CHE) model proposed in Pustejovsky and Tipton (2022) extends the three-level model by allowing estimates from the same study to have correlated estimation errors, i.e. $Cov(\epsilon_{ij}, \epsilon_{ih}) = \rho v_i^2$, where ρ is assumed to be a constant and common correlation coefficient between estimates from the same study i and $v_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij}^2$.

Table 24: Correlated Hierarchical Effects Model

ρ	Disadvantageous Inequality (α)		Advantageous Inequality (β)	
0.00	0.533	0.434	0.326	0.337
	(0.110)	(0.093)	(0.036)	(0.033)
0.10	0.531	0.432	0.328	0.339
	(0.110)	(0.092)	(0.035)	(0.032)
0.20	0.530	0.429	0.329	0.341
	(0.109)	(0.092)	(0.035)	(0.032)
0.30	0.528	0.427	0.330	0.342
	(0.109)	(0.092)	(0.035)	(0.032)
0.40	0.527	0.425	0.331	0.343
	(0.109)	(0.091)	(0.035)	(0.032)
0.50	0.525	0.422	0.332	0.345
	(0.109)	(0.091)	(0.035)	(0.032)
0.60	0.524	0.420	0.333	0.346
	(0.108)	(0.091)	(0.035)	(0.032)
0.70	0.522	0.418	0.334	0.346
	(0.108)	(0.090)	(0.035)	(0.032)
0.80	0.520	0.415	0.335	0.347
	(0.108)	(0.090)	(0.035)	(0.032)
0.90	0.519	0.413	0.336	0.348
	(0.108)	(0.090)	(0.035)	(0.032)
0.99	0.517	0.411	0.337	0.349
	(0.107)	(0.090)	(0.035)	(0.032)
Observations	149	113	144	106
Model	Full	Restricted	Full	Restricted

Notes: SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010). In column (2) the study by Diaz et al. (2023) was removed because $|DFBETAS| > 1$. In columns (3) and (4) the study by Bellemare et al. (2008) was removed because $|DFBETAS| > 1$.

G Non-Strategic versus Strategic Environments

Table 25: Meta-Analytic Averages in Non-Strategic Enviroments

	α		β	
	(1)	(2)	(3)	(4)
Constant	0.259	0.300	0.393	0.388
	(0.094)	(0.138)	(0.036)	(0.045)
p-value	0.014	0.054	< 0.0001	< 0.0001
I^2_{within}	6.73	0.57	21.92	9.85
$I^2_{between}$	93.20	99.31	77.32	88.74
Observations	65	37	84	52
Sample	Full	Restricted	Full	Restricted

Notes: All columns estimate a three-level random-effects model with the restricted maximum likelihood method; columns (2) and (4) focus on studies with reported (i.e., non-approximated) SEs; p-values are for a two-sided test with null hypothesis H_0 : Constant $= 0$; SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010).

Table 26: Meta-Analytic Averages in Strategic Enviroments

	α		β	
	(1)	(2)	(3)	(4)
Constant	0.731	0.510	0.216	0.261
	(0.169)	(0.120)	(0.064)	(0.032)
p-value	< 0.0001	< 0.0001	0.005	< 0.0001
I^2_{within}	10.21	20.89	48.10	68.88
$I^2_{between}$	89.79	79.10	51.37	28.96
Observations	84	76	60	54
Sample	Full	Restricted	Full	Restricted

Notes: All columns estimate a three-level random-effects model with the restricted maximum likelihood method; columns (2) and (4) focus on studies with reported (i.e., non-approximated) SEs; p-values are for a two-sided test with null hypothesis H_0 : Constant $= 0$; SEs in parenthesis are cluster-robust (Hedges, Tipton and Johnson, 2010). In column (2) the study by Diaz et al. (2023) was removed because $|DFBETAS| > 1$. In columns (3) and (4) the study by Bellemare et al. (2008) was removed because $|DFBETAS| > 1$.

H Ultimatum Game versus Other Games

Table 27: Ultimatum Game vs Non-Strategic and Other Games vs Non-Strategic

	Disadvantageous Inequality (α)				Advantageous Inequality (β)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ultimatum Game	0.874***	0.917***			−0.566*	−0.078**		
	(0.268)	(0.286)			(0.315)	(0.036)		
Other Games			0.091	0.308			−0.125**	−0.146**
			(0.196)	(0.269)			(0.048)	(0.055)
Constant	0.263**	0.263**	0.262***	0.262**	0.385***	0.393***	0.393***	0.393***
	(0.096)	(0.096)	(0.094)	(0.095)	(0.039)	(0.036)	(0.037)	(0.037)
I^2_{within}	18.31	18.11	3.73	3.94	44.72	21.92	31.00	27.68
$I^2_{between}$	81.65	81.86	96.25	96.02	54.81	77.31	68.02	71.45
Observations	92	90	122	78	87	85	141	97
Sample	All	Restricted	All	Restricted	All	Restricted	All	Restricted

Notes: Columns 1, 2, 5 and 6 compare estimates obtained with choices in individual decision-making tasks with estimates obtained with choices in ultimatum games. Columns 3, 4, 7 and 8 compare estimates obtained with choices in individual decision-making tasks with estimates obtained with choices in games other than the ultimatum game. Table 15 in Appendix C reports the list of games. Restricted samples in columns 2, 4, 6 and 8 exclude studies where estimates are obtained with a combination of choices in both individual decision-making tasks and games or with a combination of choices in both the ultimatum game and other games.

Table 28: Ultimatum Games vs Other Games

	Disadvantageous Inequality (α)		Advantageous Inequality (β)	
	(1)	(2)	(3)	(4)
Ultimatum Game	0.790**	0.613	−0.497	0.056
	(0.303)	(0.370)	(0.302)	(0.036)
Constant	0.352**	0.574**	0.263***	0.259***
	(0.169)	(0.246)	(0.032)	(0.036)
I^2_{within}	12.61	33.41	80.35	97.77
$I^2_{between}$	87.39	66.55	18.81	0.00
Observations	84	38	60	14
Sample	All	Restricted	All	Restricted

Notes: Columns 1, and 3 and compare estimates obtained with choices in ultimatum games with estimates obtained with choices in other games. Restricted samples in columns 2 and 4 exclude studies where estimates are obtained with a combination of choices in both individual decision-making tasks and games or with a combination of choices in both the ultimatum game and other games.

I Predictive Power of Meta-Analytic Estimates

A key contribution of our paper is to offer representative-agent estimates for α and β that serve as practical empirical benchmarks, providing a foundation for both interpretation and prediction in different settings. To evaluate the predictive ability of our meta-analytic estimates and the advancement with respect to existing estimates, we conducted a structural estimation exercise to compare their goodness-of-fit to that of the four-type distribution proposed in Fehr and Schmidt (1999), and the two-type distribution proposed in Fehr, Klein and Schmidt (2007).

For this out-of-sample test, we identified an ideal candidate in the study by Fehr, Epper and Senn (2024). Using their dataset offers several advantages: it provides a comprehensive replication package with individual-level data, includes a large and diverse sample ($N =$ 815, broadly representative of the general Swiss population), and uses tasks suitable for predictions based on models of distributional preferences. We then compared the predictive performance of the following four models on this dataset:

1. **Our Representative-Agent Model without Task Heterogeneity** defined by two parameters based on our meta-analytic estimates on the full sample ($\alpha = 0.533$ and $\beta = 0.326$).

2. **Our Representative-Agent Model with Task Heterogeneity** defined by two parameters based on our meta-analytic estimates on the non-strategic subsample ($\alpha =$ 0.259 and $\beta = 0.393$). Since the task in Fehr, Epper and Senn (2024) is an individual decision-making task, these estimates are the most appropriate.

3. **Two-Type Finite Mixture Model from Fehr, Klein and Schmidt (2007)** characterised by five parameters (a pair of parameters for each type, plus one additional parameter defining the probability of each preference type). This model assumes that α takes values of 0 or 2 with respective shares of 60% and 40%; and that β takes values of 0 or 0.6 with the same distribution.

4. **Four-Type Finite Mixture Model from Fehr and Schmidt (1999)** characterised by 11 parameters (a pair of parameters for each type, plus three additional parameters defining the probability of each preference type). This model assumes that α takes values of 0, 0.5, 1 and 4 with respective shares of 30%, 30%, 30% and 10%; and that β takes values of 0, 0.25, 0.6, and 0.6 with the same distribution.

To assess model performance, we adopt a random-utility framework with a noise component following a type-1 extreme value distribution, and evaluate each model's fit using the negative log-likelihood (NLL) and Bayesian Information Criterion (BIC). Below, we briefly describe the experimental task and data from Fehr, Epper and Senn (2024) before explaining our econometric strategy in detail.

In their experiment, participants made allocation decisions across 14 choice situations. In each situation, they chose among seven possible distributions of experimental currency units between themselves and an anonymous other participant. These seven allocations were positioned along a budget line, with the slope of the budget line varying across the 14 tasks—some positively sloped, others negatively sloped.

We model choices with a random-utility framework. Given parameters $\theta = (\alpha, \beta)$, the utility of individual i choosing allocation j in choice situation t is

$$\mathcal{U}_{ijt} \;=\; U(s_{jt}, o_{jt} \mid \theta) \;+\; \varepsilon_{ijt}$$

where $U(s_{jt}, o_{jt} \mid \theta)$ is the deterministic Fehr-Schmidt utility for allocation j in situation t, with own payoff s_{jt} and other's payoff o_{jt}, and ε_{ijt} is a noise component following a type-1 extreme value distribution with scale parameter $1/\sigma$. Let $\mathcal{X} = \{(s_{jt}, o_{jt})\}_{j=1,\ldots,7}^{t=1,\ldots,14}$ denote the full menu structure across all choice situations. With seven allocations per situation (indexed by $j \in \{1, 2, \ldots, 7\}$), this yields multinomial logit choice probabilities. Let $C_{it} \in \{1, \ldots, 7\}$ denote individual i's choice in situation t. Then

$$\Pr(C_{it} = j \mid \theta, \sigma, \mathcal{X}) = \frac{\exp(\sigma\, U(s_{jt}, o_{jt} \mid \theta))}{\sum_{k=1}^{7} \exp(\sigma\, U(s_{kt}, o_{kt} \mid \theta))}$$

Let $\mathbb{1}(C_{it} = j)$ equal one if subject i chose allocation j in situation t and zero otherwise. Subject i's likelihood contribution across the 14 choice situations is

$$f_i(\theta, \sigma \mid \mathcal{X}, C_i) = \prod_{t=1}^{14} \prod_{j=1}^{7} \Pr(C_{it} = j \mid \theta, \sigma, \mathcal{X})^{\mathbb{1}(C_{it}=j)}$$

where $C_i = (C_{i1}, \ldots, C_{i14})$ denotes individual i's full choice vector.

For finite mixture models with K latent types, each type k has parameters $\theta_k = (\alpha_k, \beta_k)$ and population share π_k (where $\sum_{k=1}^{K} \pi_k = 1$). Individual i's likelihood under the mixture model is

$$\ell_i(\Psi \mid \mathcal{X}, C_i) = \sum_{k=1}^{K} \pi_k\, f_i(\theta_k, \sigma \mid \mathcal{X}, C_i), \qquad \Psi = \big\{(\theta_k, \sigma_k)_{k=1}^{K}, (\pi_k)_{k=1}^{K}\big\}$$

Our out-of-sample exercise requires specifying the scale parameter σ. Rather than imposing an arbitrary value, we estimate σ for each model separately via maximum likelihood,

Table 29: Goodness of Fit, Representative-Agent vs Types Mixture Model

Type of Estimates	**NLL**	**BIC**
Representative-Agent Model without Task Heterogeneity	2885	5798
Representative-Agent Model with Task Heterogeneity	2854	5737
Two-Type Model from Fehr et al. (2007)	2881	5818
Four-Type Model from Fehr and Schmidt (1999)	2859	5830

Notes: NLL stands for negative loglikelihood; BIC stands for Bayesian Information Criterion.

holding the preference parameters fixed, allowing each model to achieve its best possible fit.

We evaluate model performance using two criteria: the negative log-likelihood (NLL) and the Bayesian Information Criterion (BIC). For representative-agent models with I individuals, the total NLL is

$$\text{NLL} \;=\; -\sum_{i=1}^{I} \log f_i(\theta, \sigma \mid \mathcal{X}, C_i)$$

For a K-type mixture model, it is

$$\text{NLL} \;=\; -\sum_{i=1}^{I} \log \ell_i(\Psi \mid \mathcal{X}, C_i) \;=\; -\sum_{i=1}^{I} \log\left(\sum_{k=1}^{K} \pi_k \, f_i(\theta_k, \sigma \mid \mathcal{X}, C_i)\right)$$

The BIC is defined as

$$\text{BIC} = 2 \cdot \text{NLL} + p \cdot \log N$$

where p is the number of parameters and N is the sample size. BIC augments the NLL with a complexity penalty: additional parameters must improve the fit enough to offset the $p \log N$ term. As with the NLL, lower BIC values are preferred.

Table 29 presents the model comparison results. Both representative-agent models, with and without task heterogeneity, achieve a better fit than the mixture models according to the BIC, which rewards our models' parsimony. In addition, the meta-synthetic estimates that take into account the nature of the task provide the best fit according to both criteria.

While this analysis is a proof of concept based on a single dataset, it shows that our systematic synthesis of the evidence accumulated since the introduction of the Fehr–Schmidt model has empirical value for guiding applied research in behavioral economics.

J Boxplots of Distributional Preferences Estimates

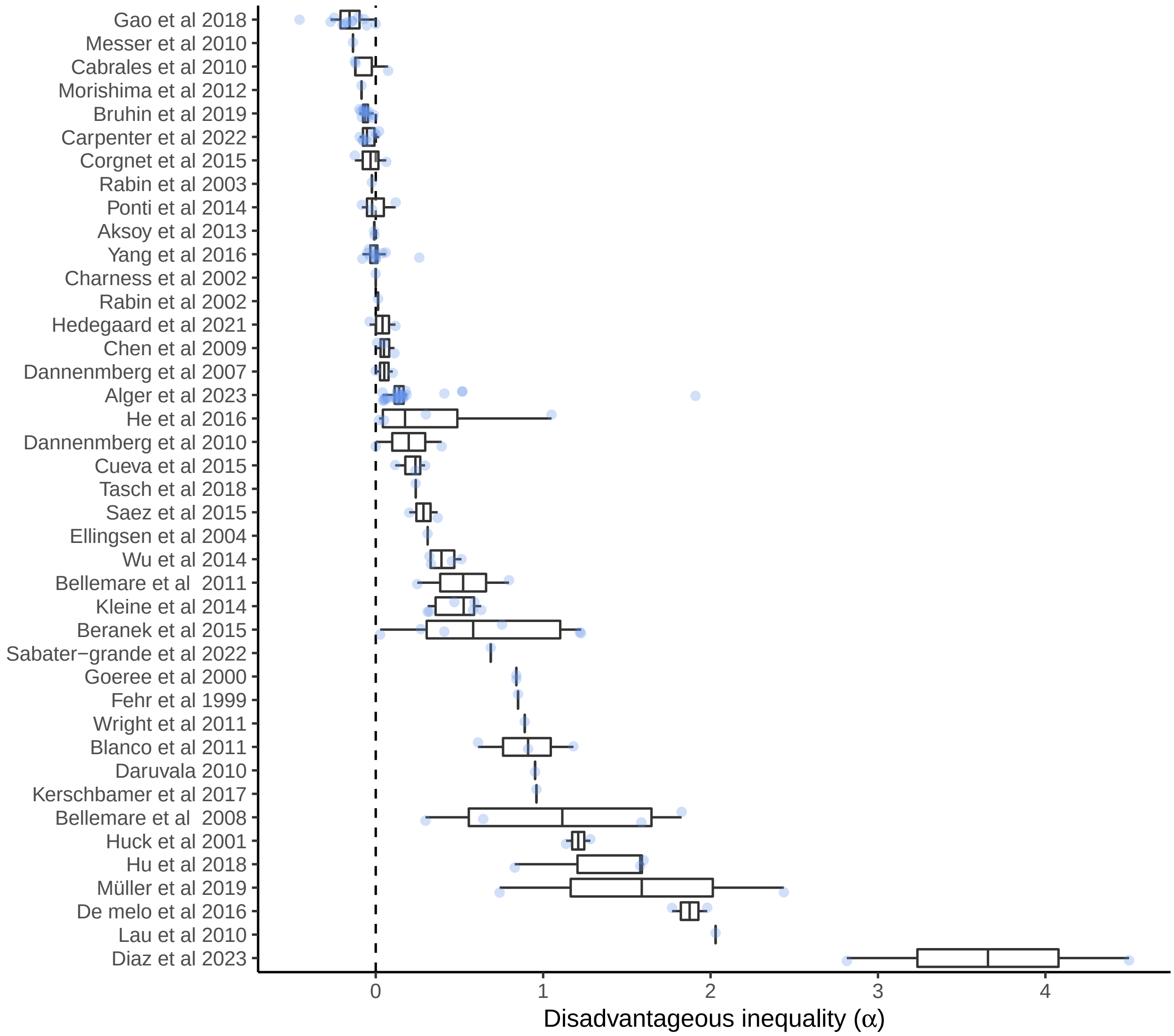


Figure 14: Boxplots of sensitivity to disadvantageous inequality estimates (α) by paper.

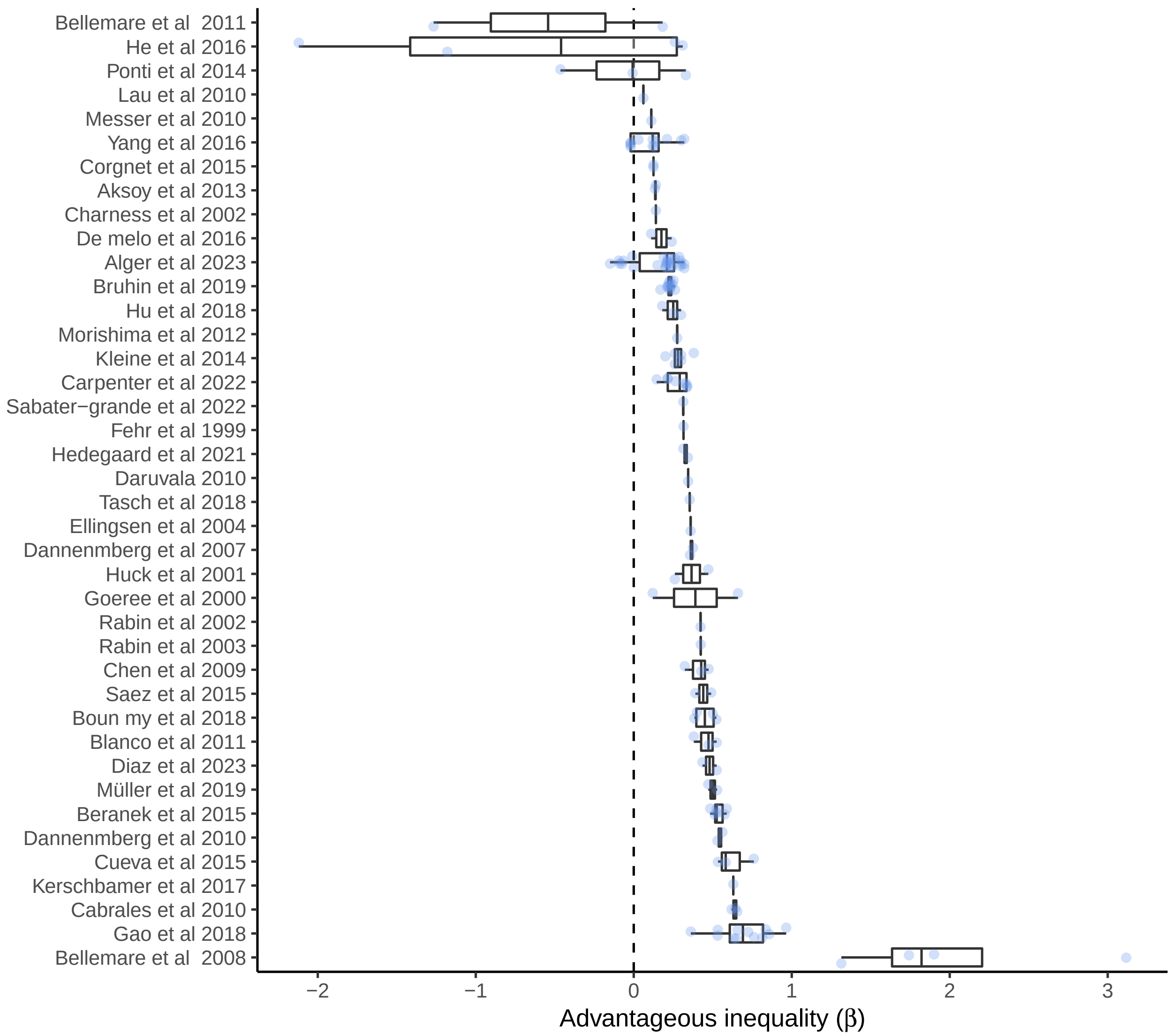


Figure 15: Boxplots of sensitivity to advantageous inequality estimates (β) by paper.

K Technical Appendix for FS Estimates

K.1 Variations of the FS Model in our Dataset

Most studies in our dataset estimate α and β assuming the utility function specification in FS. However, some studies explore variations of the original framework. First, for the sake of parsimony and mathematical tractability, FS assumed a piece-wise linear utility function. This predicts corner solutions in decision environments where we usually observe interior choices.[28] To improve on this, Bellemare, Kröger and van Soest (2008) assume a non-linear disutility from inequality and estimate the following utility function:

$$U_i(x) = x_i - \alpha_{1i} \max[x_j - x_i, 0] - \alpha_{2i} \max[x_j - x_i, 0]^2 - \beta_{1i} \max[x_i - x_j, 0] - \beta_{2i} \max[x_i - x_j]^2$$

If $\alpha_{2i} = \beta_{2i} = 0$, this model simplifies to FS. Bellemare and coauthors find the sensitivity to advantageous inequality to be nearly linear, while the sensitivity to disadvantageous inequality to be an increasing and concave function of the gap in outcomes.

A second simplification of the original model is the lack of any role for reciprocal motives. Morishima, Schunk, Bruhin, Ruff and Fehr (2012) and Bruhin, Fehr and Schunk (2019) augment FS to incorporate reciprocity, adopting the following utility function inspired by Fehr and Schmidt (1999) and Charness and Rabin (2002):

$$U_i(x_i, x_j) = (1 - \beta r - \alpha s - \theta q + \delta v)x_i + (\beta r + \alpha s + \theta q - \delta v)x_j,$$

where r, s, q, v are indicators for advantageous inequality, disadvantageous inequality, positive reciprocity and negative reciprocity respectively. Here, α and β are inequality sensitivity parameters while θ and δ are reciprocity parameters. For example, if $\theta > 0$ and $\delta < 0$, an agent rewards kind actions at a cost (i.e., he displays positive reciprocity) and punishes selfish actions at a cost (i.e., he displays negative reciprocity). Note that, in this model, the sign of the disadvantageous inequality coefficient has the opposite meaning compared to the standard FS model: here, inequity aversion is captured by $\alpha < 0$ and $\beta > 0$.[29] Bellemare, Kröger and van Soest (2011) follow another route to introduce reciprocity in FS and assume the following utility function:

$$U_i(x_i, x_j) = x_i - (\alpha_i + l_i) \max[x_j - x_i, 0] - (\beta_i + k_i) \max[x_i - x_j, 0]$$

[28] Consider, for example, a dictator game. If $\beta < 0.5$, the dictator keeps the whole budget; if $\beta > 0.5$, instead, the dictator shares the budget equally.

[29] We take this into account when using the estimates from these papers in our meta-analysis.

Table 30: FS Models in our Dataset - Estimates ($N = 297$)

	α ($N = 149$)		β ($N = 148$)	
	Frequency	Fraction	Frequency	Fraction
Utility Function in Estimated Model				
Linear FS	106	0.71	103	0.70
Non-Linear FS	4	0.03	4	0.03
Linear FS + Reciprocity	17	0.12	17	0.12
Linear FS + Kantian Morality	15	0.10	15	0.10
Linear FS + Kantian Morality + Reciprocity	5	0.03	5	0.03
Linear FS + Intentions	2	0.01	2	0.01
Linear FS + Loss Aversion	0	0.00	2	0.01

Here, depending on the intentions of the other players, l_i and k_i change the marginal disutility of disadvantageous or advantageous allocations.

Finally, the baseline FS model is sufficiently tractable to easily incorporate concerns in addition to or different from inequality sensitivity or reciprocity. For example, Alger and van Leeuwen (2024) augment the model by adding Kantian morality, whereby an individual evaluates her actions by considering what her payoff would be if others behaved in the same way; and Boun My, Lampach, Lefebvre and Magnani (2018) estimate a model of advantageous inequality aversion which includes loss aversion.

K.2 Paper Search

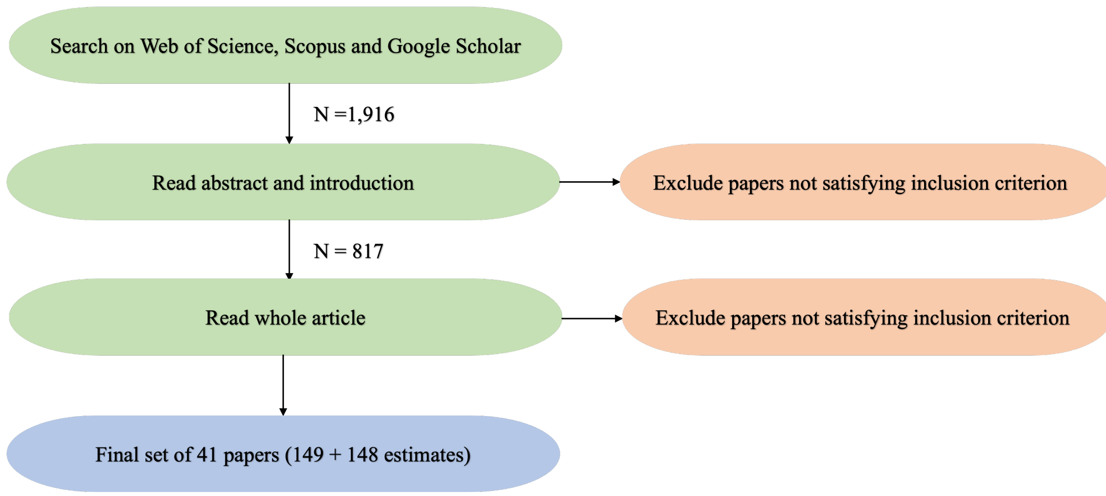
To construct our dataset, we searched relevant papers on Web of Science (February 8, 2022), Google Scholar (February 8, 2022) and Scopus (7 September 2022). We used the query in Figure 3 and looked for papers that directly cite Fehr and Schmidt (1999). This last search returned 5,665 papers citing Fehr and Schmidt (1999) and 1,916 articles that also satisfy our Query. We then read these articles and threw out 1099 papers that were clearly not relevant for our analysis. Finally, We read the remaining papers and applied our inclusion criterion—include all papers that estimated the parameters for sensitivity to disadvantageous inequality, α, and/or advantageous inequality, β, using the model by Fehr and Schmidt (1999). This last search returned the 41 articles included in our dataset.

K.3 Data Construction

We provide here more details regarding the data construction process:

- When a precise measure of the estimated parameters was not available (e.g., because the article reported only a scatter plot or a bar chart of individual-level estimates),

Figure 16: PRISMA diagram of the search



we contacted the authors to get additional details. This procedure led us to exclude a single study which computes individual-level estimates for α and β but reports only a bubble plot of these estimates (Teyssier, 2012). While it would be possible to recover an imprecise mean or median for the estimates in this study, given the high level of arbitrariness this exercise would entail (for example, in evaluating the exact location of bubbles in the graph and their relative size), we decided not to include this paper in the dataset.

- In the main body of the paper we specify how aggregate, mean, and median estimates are ready to be used in the meta-analysis. The 2 estimates from Corgnet, Espín and Hernán-González (2015) and 2 out of 4 estimates from Hedegaard, Kerschbamer, Müller and Tyran (2021) are an exception: they report set-valued individual-level estimates and the frequency of individuals in each set. In this case, we identify the interval where the median individual is located and we approximate the median value of the parameter with the median point of this interval. For example, consider an hypothetical study which estimates 6 participants have $\alpha \in [0, 0.2)$, 4 participants have $\alpha \in [0.2, 0.4)$, and 4 participants have $\alpha \in [0.4, 0.8]$. In this case, the median individual has $\alpha \in [0.2, 0.4)$ and we approximate the median individual-level estimate with 0.3.

- For the finite-mixture estimates, we computed and coded a weighted average for each parameter. For example, consider one of the finite-mixture estimates of α from Bruhin, Fehr and Schunk (2019) which reports the presence of three types in the population: $\alpha_1 = -0.159$, $\alpha_2 = -0.065$, and $\alpha_3 = 0.437$. The estimated frequencies associated with each of these types are $p_1 = 0.405$, $p_2 = 0.474$, and $p_3 = 0.121$. We construct a single estimate which is given by $\widehat{\alpha} = p_1\alpha_1 + p_2\alpha_2 + p_3\alpha_3 = -0.042$. Moreover, we construct a measure of estimation uncertainty as follows: first, we compute the standard deviation as $\text{SD} = \sqrt{\sum_i p_i(\alpha_i - \widehat{\alpha})^2}$; second, we compute the standard error as $\text{SD}/\sqrt{n}$, where n is the sample size. This procedure disregards the estimated uncertainty of each α_i and the associated p_i but it greatly simplifies our analysis and it is similar to the procedure used by studies that report an individual-level mean.

- Some studies use a combination of games and non strategic tasks to elicit the parameters. In this case, we labelled the parameters to come from a strategic environment if the number of observations obtained from games is higher than the number of observations obtained from individual decision-making tasks. The four papers are: Charness and Rabin (2002), Chen and Li (2009), Morishima et al. (2012), Bruhin et al. (2019). As for all of them the number of observations obtained in games is higher, they are labelled as estimates coming from strategic environments.

K.4 Approximation of Standard Errors

Out of 297 estimates in our dataset, the source reported the SEs for 79 estimates and, in other 146 cases, we were able to compute the SEs using the reported standard deviation and sample size or from t-statistics. For the remaining 72 estimates, we did not have (direct or indirect) information about the SEs. We had two options: either drop the 72 estimates without SEs or approximate the SEs and keep these estimates in the dataset. We chose the latter option, especially since the observations would not be dropped randomly: as the density plots in the top row of Figure 17 show, there is a difference in the distribution of α and β between studies that report SEs and studies that did not and, thus, dropping the latter subset of estimates would introduce a bias in our results.[30] For this reason, while using approximated SEs is a second-best, we deemed this as the more sensible option.

For approximating unavailable SEs, we followed the procedure in Brown, Imai, Vieider and Camerer (2024): we first estimated the parameters characterizing the distribution in the data as $log(se_o) \sim \mathcal{N}(\mu_{se}, \sigma^2_{se})$; and we then used these distributional parameters to estimate the missing SEs as $log(se_m) \sim \mathcal{N}(\hat{\mu}_{se}, \hat{\sigma}^2_{se})$, where o stands for observed and m stands for missing. In order for this procedure to give a good approximation of the SEs, we need variables that are significantly associated with them. In our dataset, the values of the parameters are the best predictors for the values of their SEs. Other information available to us does not improve the estimates.[31] We, thus, run the two following regressions to find

[30] The two distributions of β are statistically different according to a Wilcoxon rank sum test.

[31] This information includes whether the study is conducted in the laboratory, in the classroom or online;

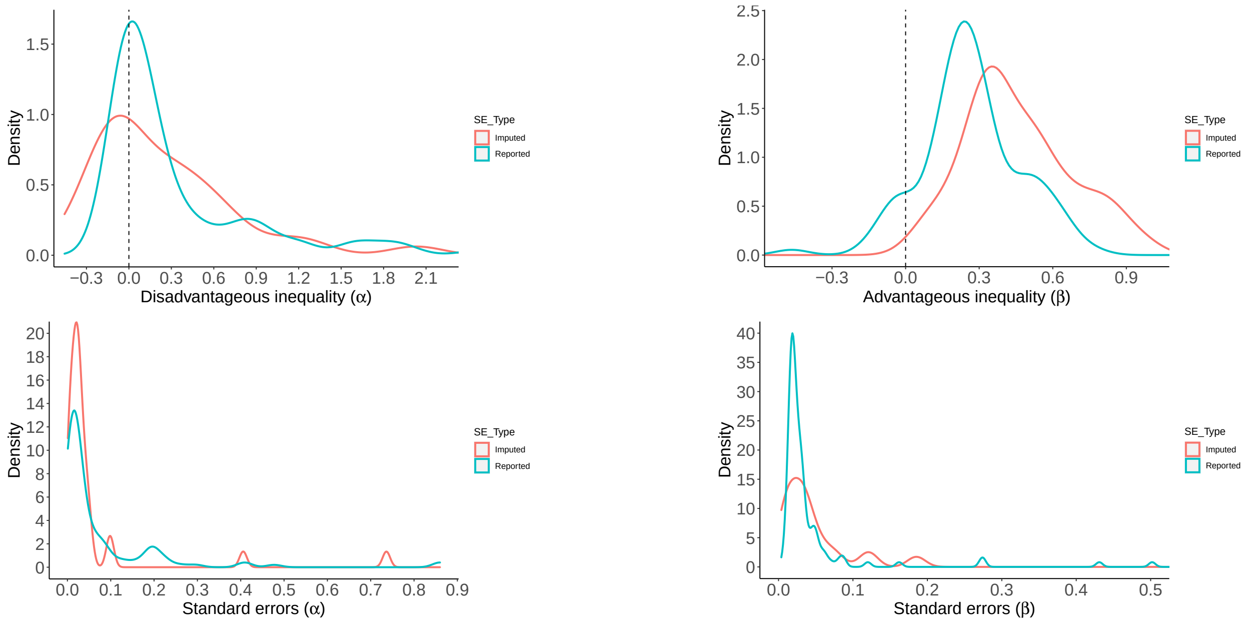


Figure 17: Distribution of Estimates and SEs for α and β as Function of SE Type. Note: The top two graphs show kernel density estimates (Gaussian with Silverman's rule of thumb) for the subsets of parameters with reported vs. imputed SEs; the bottom two graphs show kernel density estimates of SEs in the two subgroups; the x-axis in the density plot for α is truncated at 2.2; the x-axis in the density plot for β is truncated at -0.5 and 0.9 for better visual rendering but the kernel density uses all estimates in both cases.

$\hat{\mu}^{\alpha}_{se}$, $\hat{\mu}^{\beta}_{se}$ and their respective variances:[32]

$$log(se^{\alpha}_{o}) = \delta_0 + \delta_1 \alpha_o + \delta_2 \beta_o$$

$$log(se^{\beta}_{o}) = \gamma_0 + \gamma_1 \alpha_o + \gamma_2 \beta_o$$

whether subjects are college students, a convenience sample of adults or a representative sample of the general population; whether the estimate is an individual-level mean, an individual-level median, an aggregate mean or from a finite mixture model; and what version of the FS model was estimated.

[32]For studies estimating a single parameter, we use only this estimate (and a constant) as regressor.

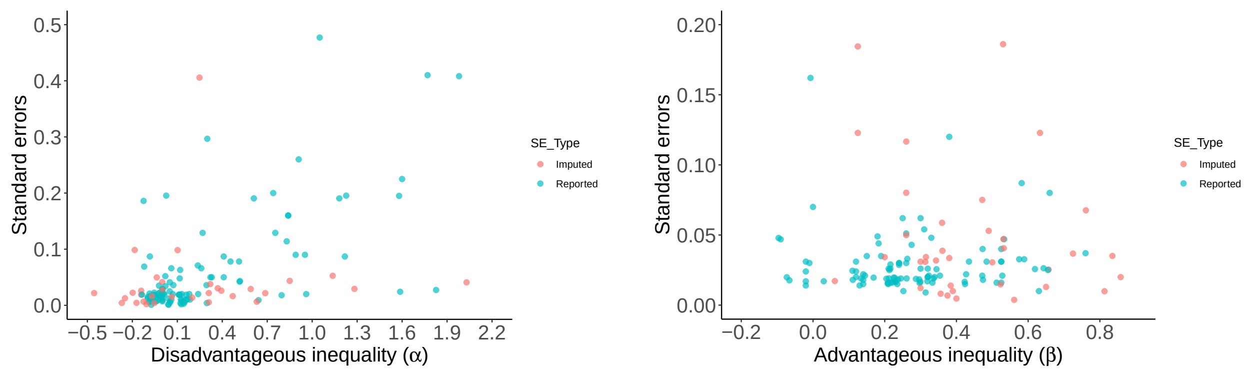


Figure 18: Scatter Plots of α and β SEs as a Function of SE Type. Note: The x-axis in the plot for α is truncated at 2.2; the x-axis in the plot for β is truncated at -0.2 and 0.9 for better visual rendering.

The two parameters explain 35% of the variance in the SEs for α and 11% of the variance in the SEs for β. Our approximation is, thus, better for α than for β.

L Technical Appendix for AM Estimates

L.1 Articles Included in Dataset (Chronological Order)

1. **Andreoni, James, and John Miller**, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 2002, 70, no. 2: 737-753.

2. **Andreoni, James**, "Giving Gifts to Groups: How Altruism Depends on the Number of Recipients," *Journal of Public Economics*, 2007, 91, no. 9: 1731-1749.

3. **Fisman, Raymond, Shachar Kariv, and Daniel Markovits**, "Individual Preferences for Giving," *American Economic Review*, 2007, 97, no. 5: 1858-1876.

4. **Fisman, Raymond, Pamela Jakiela, and Shachar Kariv**, "How did Distributional Preferences Change During the Great Recession?," *Journal of Public Economics*, 2015, 128: 84-95.

5. **Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits**, "The Distributional Preferences of an Elite," *Science*, 2015, 349, no. 6254: aab0096.

6. **Fisman, Raymond, Pamela Jakiela, and Shachar Kariv**, "Distributional Preferences and Political Behavior," *Journal of Public Economics*, 2017, 155:1-10.

7. **Li, Jing, William H. Dow, and Shachar Kariv**, "Social Preferences of Future Physicians," *Proceedings of the National Academy of Sciences*, 2017, 114, no. 48: E10291-E10300.

8. **Li, Jing**, "Plastic Surgery or Primary Care? Altruistic Preferences and Expected Specialty Choice of US Medical Students," *Journal of Health Economics*, 2018, 62: 45-59.

9. **Lopez-Persem, Alizée, Lionel Rigoux, Sacha Bourgeois-Gironde, Jean Daunizeau, and Mathias Pessiglione**, "Choose, Rate or Squeeze: Comparison of Economic Value Functions Elicited by Different Behavioral Tasks," *PLoS computational biology*, 2017, 13, no. 11: e1005848.

10. **Müller, Daniel**, "The Anatomy of Distributional Preferences with Group Identity," *Journal of Economic Behavior & Organization*, 2019, 166: 785-807.

11. **Breitmoser, Yves, and Jonathan HW Tan**, "Why Should Majority Voting be Unfair?," *Journal of Economic Behavior & Organization*, 2020, 175: 281-295.

12. **Flora Li, Sheryl Ball, Xiaomeng Zhang, and Alec Smith**, "Focal Stimulation of the Temporoparietal Junction Improves Rationality in Prosocial Decision-Making," *Scientific Reports*, 2020, 10, no. 1: 20275.

13. **Robson, Matthew**, "Inequality Aversion, Self-Interest and Social Connectedness," *Journal of Economic Behavior & Organization*, 2021, 183: 744-772.

14. **Erkut, Hande**, "Social Norms and Preferences for Generosity are Domain Dependent," *Games and Economic Behavior*, 2022, 131: 121-140.

15. **Li, Jing, Lawrence P. Casalino, Raymond Fisman, Shachar Kariv, and Daniel Markovits**, "Experimental Evidence of Physician Social Preferences," *Proceedings of the National Academy of Sciences*, 2022, 119, no. 28: e2112726119.

16. **Attema, Arthur E., Matteo M. Galizzi, Mona Gross, Heike Hennig-Schmidt, Yassin Karay, Olivier L'haridon, and Daniel Wiesen**, "The Formation of Physician Altruism," *Journal of Health Economics*, 2023, 87: 102716.

17. **Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Silvia Vannutelli**, "The distributional preferences of Americans, 2013–2016," *Experimental Economics*, 2023, 1-22.

L.2 Paper Search

As mentioned in the main body of the paper, the inclusion criteria was to include "all studies that estimate the parameters for the weight on own payoff, a, and/or equality/efficiency, ρ." The search was made on Scopus (July 3, 2023), and we looked at papers that cited Andreoni and Miller (2002) and contained the word "Estimat*" and "Elasticity." The inclusion of "Estimat*" in the Query is straightforward given our objective. We also add the word "Elasticity" as we believe it is a must use word for studies estimating the parameters of Andreoni and Miller (2002), as it can come up in the description of the constant elasticity of substitution utility, the meaning of the parameter for equality/efficiency ρ, or the interpretation of the results obtained. This search returned 68 articles, including Andreoni and Miller (2002). We then read all the articles satisfying the Query and applied the inclusion criteria. The final dataset consists of 18 studies from 17 papers, and 98 estimates, 49 for a and 49 for ρ. The paper containing two studies is Fisman et al. (2015), with a sample of Yale Law students and Elite ALP (American Life Panel) subjects.

Table 31: Features of the Estimates ($N = 98$) in the Dataset.

	a ($N = 49$)		ρ ($N = 49$)	
	Frequency	Fraction	Frequency	Fraction
Type of Estimates				
Aggregate	7	0.14	7	0.14
Finite Mixture	11	0.22	11	0.22
Individual Mean	19	0.39	19	0.39
Individual Median	12	0.25	12	0.25
Estimation Method				
Tobit	30	0.61	30	0.61
Logit	7	0.14	7	0.14
Other	12	0.25	12	0.25
Standard Errors				
Reported	46	0.94	46	0.94
Imputed	3	0.06	3	0.06

L.3 Data Construction and Feature of Estimates in the Dataset

For 46 estimates out of 49 for a and ρ we have information about the SEs, or can compute them with information inside the articles. For three estimates from one paper we do not have the SEs. In this case we followed the same procedure we did for the estimates from the FS model, described in Section K.4.

Table 31 presents the features of the 49 estimates for a and 49 estimates for ρ included in our dataset. The majority of the estimates, around 65%, come from studies reporting the mean or median value of the parameters, while 14% and 22% come from studies reporting aggregate or finite mixture values of the parameters. The most common estimation method consists in a Tobit procedure on the demand function of the modified dictator game. One study employing the mini-dictator game with equality-efficiency tradeoffs and one study employing a majority bargaining game use instead a logit framework. The remaining estimation methods consist of maximum likelihood on the demand function of the modified dictator game with normal noise, or employing a Dirichlet distribution, or using Variational Bayesian Analysis on the choices made from a mini-dictator game with equality-efficiency tradeoffs.