For the latest version, please click here
For the Supplementary material, please click here

Redistribution Over Gains and Losses: Social Preferences and Moral Rules

Michalis Drouvelis

University of Birmingham

Ernesto M. Gavassa-Pérez

University of Navarra

December 5, 2024

We investigate redistributive behavior over gains and losses. Using two pre-registered experiments, we document a systematic asymmetry in behavior: people are more selfish when redistributing over losses than over equivalent gains. We use structural estimation methods and out-of-sample predictions to understand the drivers of choices made by experimental subjects, and identify that a mix of social preferences coupled with loss aversion (inequality aversion, social efficiency, and maximin) alongside moral rules (blame avoidance and praise seeking) are key to understand the individual heterogeneity of redistributive behavior. *JEL Codes*: C79, C91, D63, D91.
Keywords: Dictator Games, Social Preferences, Disinterested Morality, Structural Estimation.

I. Introduction

Other-regarding behavior is pervasive in human life, and throughout the last half a century behavioral economics has documented extensively altruistic departures from selfish motives. The earliest theoretical rationale to

. **Drouvelis** (*Corresponding Author*): University of Birmingham, The Department of Economics, University House, Edgbaston, B15 2TT, Birmingham, UK (e-mail: m.drouvelis@bham.ac.uk); **Gavassa-Pérez**: University of Navarra, Office 2020, Amigos Building, Calle Universidad, 1, 31009, Pamplona, Navarra, Spain (e-mail: egavass@unav.es). We want to express special gratitude to Jose Apesteguia and Markus Kinateder for detailed comments on the draft of the paper and to Nina Serdarevic for her input at an earlier stage of the idea conception. We'd like to thank seminar participants of the 2023 CCC (Nottingham) meeting, ESA Europe (Exeter, 2023) conference, Behavioral & Experimental Economics Workshop (Navarra, 2024), Newcastle Experimental Economics Workshop (Newcastle, 2024) and the 94th Southern Economics Association (Washington DC, 2024) conference for useful comments on the research enclosed in this manuscript.

predict social behavior was altruism as presented by Becker (1974) . Despite the importance of those initial efforts, and driven by the experimental literature of the 70's and 80's, the models had important limitations when it came to predict economic behavior, such as charitable giving. Several discussions, such as the one in Sugden (1982), motivated further theoretical developments and experimental research. One key paper was that of Güth, Schmittberger, and Schwarze (1982), who documented how, in donation decisions where receivers had the chance to reject the donation and burn the money, people gave significant amounts to others. One explanation for giving in such situations was the role of intentions and the threat of punishing unkind actions, and reciprocity did not take long to land in economic models pioneered by Sugden (1984) and Rabin (1993). However, further evidence was gathered by Forsythe et al. (1994), who documented that, even when receivers did not have the possibility to veto the amount received, an important proportion of people donated money; which was not predicted by reciprocity.

All the experimental evidence accumulated, asking for new theoretical models that could serve as explanations for the social behavior documented in experiments; and models of distributional preferences ensued. Fehr and Schmidt (1999) proposed a model of inequality aversion, but it could not rationalise, in its linear shape, donations below 50% of one's own wealth; and further models where developed: non-linear ones, like the one in Bolton and Ockenfels (2000); models capturing concerns for the aggregate social wealth and the Rawlsian maximin, as the one in Charness and Rabin (2002); and models capturing moral motivations, such as the ones in Levitt and List (2007), Roemer (2010), and Alger and Weibull (2013) among others[1].

Amongst all the experimental games used to capture prosocial tendencies, the Dictator Game reported in Forsythe et al. (1994) is arguably the most used one. It has been used to document how social norms influence giving, such as in Andreoni and Bernheim (2009); how entitlements determine the degree of selfishness, such as in Cappelen et al. (2007) and Oxoby and Spraggon (2008); and how culturally variable prosociality is, such as in Henrich et al. (2001)[2]. In a seminal paper, List (2007) presented evidence showing that allowing to take money from the receiver in a dictator game made behavior more selfish. This evidence was replicated in Bardsley (2008) and Cappelen et al. (2013), providing distributional preferences with a new

1. See, for instance, the summaries in Camerer (2003), Konow (2003), Sobel (2005), Cooper and Kagel (2016), and Bowles and Polania-Reyes (2012).

2. For another important work on social norms and dictator game giving, see Krupka and Weber (2013). The dictator game has been used extensively to document several empirical regularities. We refer the reader to the following sources. For a summary of dictator games in developing economies, see Cardenas and Carpenter (2008). For how beliefs are important drivers of giving, see Dana, Cain, and Dawes (2006). For gender differences in dictator games, see Eckel and Grossman (2001). For other cross-cultural experiments involving dictator games, see Henrich et al. (2005). For the influence of anonimity and social distance in dictator games, see Hoffman et al. (1994) and Hoffman, McCabe, and Smith (1996), Cherry, Frykblom, and Shogren (2002), and Goeree et al. (2010). For further work on rules, fairness, and dictator game giving, see Bolton, Katok, and Zwick (1998), Konow (2003), and Schurter and Wilson (2009). Finally, for a meta-analysis surveying dictator games, see Engel (2011).

limitation: if all that matters is distributional concerns, one would expect those to be stable regardless of whether the domain of play is gains, losses, or a mixture of both[3]. Boun My et al. (2018) provide a further tweak to the experimental design of dictator games, where subjects make decisions that have identical distributional consequences, but where some are dictator games over gains and others are dictator games over losses. The authors document, similarly to what was observed by List (2007), that subjects are more selfish under the presence of losses, a behavior that has been replicated in Fiedler and Hillenbrand (2020). Although classical distributional preferences cannot account for such behavior, if one couples them with reference dependence, as in Breitmoser and Tan (2013); or if the moral cost of giving is higher under losses, in the vein of the model proposed in Levitt and List (2007), one can account for this asymmetry in behavior.

Understudied in the behavioral and experimental economics literature, the role of normative judgments can also be a motivational force driving behavior. Harsanyi (1955) coined the term *ethical preferences*, by which he referred to social wefare considerations independent from the concept of utility as traditionally understood within economics. Sen (1977) referred to a counter-preferential path to moral doing as *commitment*, where one's choice was independent from how the welfare of others affected one's utility. Roth (2007) introduced moral repugnance as constraints, rather than as a new term in someone's preference profile. And more recently, Smith and Wilson (2019) have proposed a framework based on Adam Smith's work, where the moral considerations of an impartial spectator, rather than utility, are the main drivers of behavior in the social world. Gavassa-Pérez (2022) builds on the latter to develop a new theoretical framework that blends impartial considerations with heterogeneous moral perceptions, adding a pinch of moral relativism to the equation to be able to accommodate how differing moral views can drive different behaviors. Despite the important theoretical efforts trying to separate preference from choice, the most canonical experimental papers reported that social preferences explained social behavior, but remained silent as of whether the counter-preferential approach could as well explain the same data.

Thus, experiments explicitly considering impartial moral rules as the ultimate mechanism explaining distributional choices remain yet to be published. In this study, we structurally estimate the parameters and moral judgments of several candidate theories and use them to make out-of-sample predictions in binary dictator games over gains and losses. We focus on two potential mechanisms of the asymmetric distributional behavior over gains and losses: reference-dependent social preferences and impartial moral rules. We expose each subject to a set of tasks to structurally measure their rele-

3. Another related paper exploring the link between reference dependence and dictator game giving is Benistant and Suchon (2021). More generally, for a recent summary of the research on dictator game giving and losses, see Cochard and Flage (2024). Although there exists conflicting evidence on whether dictator game giving is more selfish over losses, all the papers following the designs of List (2007) and Boun My et al. (2018) have reported more selfishness over losses.

vant parameters and impartial moral judgments, and use the data to make individual-level point predictions in 22 binary dictator games, of which 11 involve redistributing gains and 11 involve redistributing losses.

We report experimental data from two studies following a similar design. The first study comprises 305 laboratory subjects. The second study provides a replication in an online experiment recruiting 348 subjects that form a representative sample of the UK population. We document a systematic tendency at the within-subject level to engage in more selfish behavior over losses than in equivalent distributional situations over gains; which is even stronger in the representative sample. Additionally, we find (and replicate) that both contextual features, such as the size of the egalitarian payoff and whether distributional situations involve gains or losses, and motivational forces, such as reference-dependent social preferences and impartial moral rules, are important drivers of behavior.

Our results question the traditional interpretation that other-regarding preferences alone are the mechanism driving distributional choices. We demonstrate the importance of coupling reference-dependence with social preferences in order to rationalise around 50% of the giving decisions in dictator games. Moreover, even when controlling for other-regarding preferences, we find that the inclusion of impartial moral rules introduced in Gavassa-Pérez (2022) are important drivers of around 35 to 39% of giving behavior. This is important, as the theoretical framework of moral rules we use builds on the spirit of canonical work in our discipline challenging the tight link between preference, or utility, and choice[4].

The current experimental design cannot explicitly rule out that other preferences can rationalise the behavior predicted by moral rules. However, it documents that a different path of social doing, currently under-explored in the economics literature, can as well account for the social behavior that challenged the motivational structure underpinning the Homo economicus. It, then, opens the door for future experimentation to disentangle between both alternative accounts of social decision-making.

The rest of the paper is organised as follows. Section II. presents the experimental design. Section III. discusses the empirical strategy of the paper. Section IV. presents the results we find and section V. discusses the implications of the results and concludes.

II. Experimental Design

We run two experiments, one of which was implemented in the lab (henceforth, *Lab*) and the other one in a representative sample of the UK population (henceforth, *Representative Sample*). Both experiments consist of the

4. See Harsanyi (1955); Sen (1977); Roth (2007); and Smith and Wilson (2015, 2017, 2019) for important discussions on the topic.

same five sets of tasks. In this section we present all tasks and discuss the different treatments that result from presenting them in different orders. We highlight differences in the experimental implementation – there are no differences apart from those stated in the text.

A. Main Games: Dictator Games

The games that are the main focus of our investigation are modified dictator games. In the classic dictator game, presented in Forsythe et al. (1994), two persons interact in a decision situation; a proposer (*dictator* from now onwards) that can split a sum of money between them and a responder (*passive agent* from now onwards), which has to accept whatever division of payoffs the proposer chooses. In this paper we simplify the dictator games as in Blanco, Engelmann, and Normann (2011), as we give dictators only two potential divisions of money.

We use the 22 binary dictator games presented in boun et al. (2018), where the dictator has to choose between an unequal distribution of money and an egalitarian distribution of money. Out of the 22 binary dictator games, eleven involve distributions of gains and the remaining eleven involve distributions of losses. Relative to the unequal distribution, choosing the egalitarian distribution implies the dictator loses 'a' units of payoff to increase the passive agent's payoff by 'b' units. Thus, the ratio $\frac{b}{a}$, which we refer to as *efficiency* from now onwards, captures the ratio of social gains over social losses induced by implementing the egalitarian over the unequal distribution of money. Table 1 presents the 22 modified dictator games under investigation in 11 rows, where games in the same row are equivalent in terms of efficiency but differ on whether the distribution of money involves gains or losses[5].

B. Structural Estimation Tasks

1. Loss Aversion

The *Loss Aversion* tasks involved using the BDM mechanism[6] to elicit willingness to accept (henceforth, *wta*) and pay (henceforth, *wtp*) estimates for a mug in an incentive compatible manner. In the *wta* (resp.,*wtp*) task, we told subjects they were endowed with a mug (resp., £10), and that they had to provide a valuation of the mug. We stated we would randomly draw a number between £0 and £10, and that if the randomly drawn number was higher (resp. lower) than their valuation, a transaction would be made. In the *wta* task, this implied subjects would sell the mug to the experimenter for the randomly drawn price. In the *wtp* task, this implied subjects would

5. modified dictator games equivalent in terms in efficiency are also equivalent in terms of the magnitude of inequality

6. see Becker, Degroot, and Marschak (1964) for details of the experimental procedure.

TABLE 1
BINARY DICTATOR GAMES OVER GAINS AND LOSSES

	Gains		Losses	
	Unequal	Egalitarian	Unequal	Egalitarian
1	(£10, £0)	(£0, £0)	(£0, −£10)	(−£10, −£10)
2	(£10, £0)	(£1, £1)	(£0, −£10)	(−£9, −£9)
3	(£10, £0)	(£2, £2)	(£0, −£10)	(−£8, −£8)
4	(£10, £0)	(£3, £3)	(£0, −£10)	(−£7, −£7)
5	(£10, £0)	(£4, £4)	(£0, −£10)	(−£6, −£6)
6	(£10, £0)	(£5, £5)	(£0, −£10)	(−£5, −£5)
7	(£10, £0)	(£6, £6)	(£0, −£10)	(−£4, −£4)
8	(£10, £0)	(£7, £7)	(£0, −£10)	(−£3, −£3)
9	(£10, £0)	(£8, £8)	(£0, −£10)	(−£2, −£2)
10	(£10, £0)	(£9, £9)	(£0, −£10)	(−£1, −£1)
11	(£10, £0)	(£10, £10)	(£0, −£10)	(£0, £0)

buy the mug from the experimenter in exchange for the randomly drawn price, keeping the remainder of the £10 for themselves.

In the laboratory experiment, we presented students real mugs at the beginning of the experimental session and told them to inspect them if they wanted to. For logistical reasons, this was not possible to implement online. To ensure subjects could develop some sense of attachment to the good in both experiments, we provided pictures of the mug alongside some characteristics of it (see experimental instructions for more detail on the exact implementation) in the experimental screen before their tasks started.

Since the seminal Kahneman, Knetsch, and Thaler (1990) article was published, these tasks have become a workhorse within the experimental economics literature to capture disparities between willingness to pay and willingness to accept. Recently, Gächter, Johnson, and Herrmann (2022) used this task as a simple structural measure of loss aversion in the riskless domain, and we follow them in doing so.

2. Social Preferences

The *Satisfaction Ratings* tasks involve 64 different distributions of money between four persons. We tell each subject to assume the role of one of the four persons in each of the 64 distributions. For each of the 64 income distributions, we ask subjects to indicate their satisfaction on a scale from −50

(extremely dissatisfied) to +50 (extremely satisfied)[7]. The subjects have the information on the distribution at the time of self-reporting their satisfaction. We randomised the order in which we presented each distribution to subjects.

The task is inspired by Loewenstein, Thompson, and Bazerman (1989), who use the satisfaction ratings to estimate social preferences. Like them, satisfaction ratings are not incentivised. Although this takes a different approach from the literature of structural estimation of social preferences in experimental economics, such as Andreoni and Miller (2002), Fisman, Kariv, and Markovits (2007), Bellemare, Kröger, and Van Soest (2008), Iriberri and Rey-Biel (2013), and Bruhin, Fehr, and Schunk (2018), several papers have proved that self-reported measurements are valid tools to elicit social preferences (see, for instance, the original Loewenstein, Thompson, and Bazerman (1989) study, and more recently the study reported in Diaz et al. (2023)[8]).

One important design decision was whether to use the same tasks to elicit distributional concerns and loss aversion, and we opted not to do so for several reasons. First, notice that decisions in the Loss Aversion tasks are incentivised, and outcomes of them are kept private, so there are no reference points on other subject's payoffs. Hence, there is no reason to believe that distributional concerns might bias responses in the Loss Aversion tasks. Second, the Loss Aversion tasks we use have been implemented pervasively in the literature, so we believe they are a better elicitation method than the alternative of using a new method to capture the same effect. Finally, the satisfaction ratings only involve gains, so there is no reason to believe loss aversion is influencing self-reported satisfaction ratings of distributions over gains. Given all of the above, we are confident that using the Loss Aversion and Satisfaction Ratings tasks allows us to extract parameters of social preferences and loss aversion that we can couple together as estimates of models of reference-dependent social preferences.

C. Moral Judgment Scenarios

In the *Moral Judgments task*, we present several modified dictator games of the type described earlier to the experimental subjects, and tell them each game is played between a dictator (*Person A*) and a passive subject (*Person B*). In contrast with the Satisfaction Ratings, we explicitly tell them they

7. Other papers have used a likert scale from 1 to 7 when eliciting satisfaction rsatings. We opt for a more continuous scale as some research has proven that, when two variables are normal, dichotomising one of them reduces the initial correlation and generates a higher risk of false positives (see Maxwell and Delaney (1993) and McClelland et al. (2015)). Additionally, the moral judgments tasks presented later on are elicited on the same scale. Given that we use both tasks to make predictions of different theories, we wanted to keep the elicitation methods as similar as possible so that all differences in out-of-sample success are due to their underlying importance as behavioral mechanisms

8. More broadly, subjective wellbeing measures are commonly used in economic research. See, for instance, the works of Di Tella, MacCulloch, and Oswald (2001); Frey and Stutzer (2002); Blanchflower and Oswald (2004); Clark, Frijters, and Shields (2008); Benjamin et al. (2012); Card et al. (2012); and Deaton and Stone (2013) among others.

are *impartial spectators* in these games. That is, they are *neither* Person A *nor* Person B, they know nothing about the identity of the Persons involved in the dictator games, and they *don't have stakes* in the decision situation (i.e., their moral judgments are self-reported and unincentivized).

For each modified dictator game, we ask our impartial subjects to give a moral rating of Person A given the pieces of information we provide them, which are (i) the payoff structure of the game; and (ii) Person A's action. Given that we use the 22 payoff structures of the dictator games in Table 1, and that there are two possible actions for each game, this implies each subject provides 44 impartial moral ratings of Person A. The moral rating is elicited on a scale from -50 (extremely bad) to $+50$ (extremely good)[9]. We randomised the order in which we presented each of the 22 payoff structures to subjects.

D. Sociodemographic Questionnaire

All subjects answered a set of sociodemographic questions at the end of the experiment. Some questions were different in the Lab and the Representative Sample experiments. This served to elicit control variables in our regression analysis, such as gender, political affiliation, religiousness, income, age, and level of education.

E. Treatment Manipulations

Every experimental subject made choices in the 22 modified dictator games, gave their 44 impartial moral judgments, provided their willingness to accept and pay estimates, and stated their 64 self-reported satisfaction ratings. In total, each subject provided us with 132 distinct data points on top of the answers to the sociodemographic questionnaire.

To reduce the number of potential orders in which we could present the experimental tasks to subjects, we fixed the order of some tasks. First, we presented the satisfaction ratings in between the modified dictator games over gains and losses. Second, subjects always answered the sociodemographic questionnaire at the end of the experiment. Third, we presented the Loss Aversion tasks one after each other. And fourth, dictator games and moral judgments always followed the same sequence. That is, if dictator games over gains preceded dictator games over losses for a subject, then the moral judgments of dictator games over gains also preceded those over losses for them.

Subject to those restrictions, we randomised (i) whether we presented moral judgments before the dictator games; (ii) if we presented the Loss

9. The moral judgment task follows the same implementation of Cubitt et al. (2011) and Gavassa-Pérez (2022). For more details on how the moral scenarios were presented to subjects, one can refer to the experimental instructions provided in the online materials of this paper.

Aversion tasks before, or in between, the moral judgments and dictator games; (iii) whether *wta* preceded *wtp*; and (iv) whether gains precede losses in both dictator games and moral judgments. Given that each factor has two possible orderings, these four manipulations imply 16 different orderings in which we could present tasks to subjects.

F. Differences between the two experiments

We recruited 305 university students from the University of Birmingham in the first experiment, and a representative sample of the UK Population (348 subjects) from Prolific in the second experiment. The sample size was calibrated in both cases to achieve 80% statistical power on some of the statistical tests presented below[10].

Dealing with sending individual mugs to participants in the Representative Sample experiment was prohibitively challenging given that the online nature of the experiment generated shipping costs, and, more importantly, delays in the experimental payoff (if the mug was sold) not present in the Lab experiment. For that reason, we implemented the Representative Sample experiment using the Conditional Information Lottery method presented in Bardsley (2000), rather than the Random Lottery Incentive System used for the implementation of the Lab experiment. Although this introduces an asymmetry in experimental implementation, both methods are incentive-compatible, and should not systematically alter people's behavior if they act as rational agents[11].

III. Identification Strategy

A. Structural Estimation of Loss Aversion and Social Preferences

Below we present the equations we estimate to retrieve the relevant parameters. For social preferences, we present those alongside a presentation of the theoretical models we want to estimate[12].

Loss Aversion. We estimate the coefficient of loss aversion as $\hat{\lambda}_i = \frac{wta}{wtp}$.

Social Preferences. We use the elicited satisfaction ratings to make

10. See both pre-registration documents for a detailed discussion of the sample size rationale

11. In the Random Lottery, subjects are told all tasks are played for real, and that one task will be chosen at random for payment. In the Conditional Information Lottery, subjects are told that all but one of the tasks are fictional, and that they will be paid taking into account their actions in the real task. It is crucial to note that in the Conditional Information, subjects do not know what the real task is. This mimics the spirit of the Random Lottery method, in which the task is chosen after all subjects have finished the experiment. Thus, in both cases subjects have incomplete information about the task that will be chosen for payment, and hence they have incentives to reveal their preferences in every incentivised task.

12. In this section, we keep the discussion of loss aversion and social preferences separate given that we use different tasks to calibrate each set of parameters. In the appendix, we present the models merging both social preferences and loss aversion.

structural estimations of the following models of social preferences:

$$U_i\ (\pi_i, \pi_j\ ) = \pi_i - \frac{\beta_i}{n-1} \cdot \left(\sum_{j \neq i} \max\{\pi_j - \pi_i, 0\} \right) \tag{1}$$

$$U_i\ (\pi_i, \pi_j\ ) = (1-\rho_i) \cdot \pi_i + \rho_i \cdot \sum_{j=1}^{n} \pi_j \tag{2}$$

$$U_i\ (\pi_i, \pi_j\ ) = (1-\gamma_i) \cdot \pi_i + \gamma_i \cdot \min\{\pi_1, \cdots, \pi_n\} \tag{3}$$

The first model refers to a modified version of Fehr and Schmidt's (1999) model of inequality aversion, where we omit concerns for disadvantageous inequality (i.e., α_i) as it is not behaviorally relevant for dictator game giving[13]. The restrictions to the parameter space are $1 > \beta_i \geq 0$, where β_i captures a concern for advantageous inequality (i.e., having more material payoff than others). The second and third models dis-aggregate Charness and Rabin's (2002) two motivations into two models, one capturing concerns of *social efficiency* (i.e., aggregate social welfare $\sum_{j=1}^{n} \pi_j$) and another one capturing *rawlsian maximin* concerns (i.e., $\min\{\pi_1, \cdots, \pi_n\}$). In the latter two models, we impose restrictions $\rho_i, \gamma_i \in [0, 1]$ to the parameter space, where ρ_i (resp. γ_i) represent a subject's concern for the relevant social motivation.

Letting us define S_{id} as the satisfaction rating of subject i from distribution d, we can define the three structural models we estimate separately for each subject i with the following non-linear equations

$$\hat{S}_{id}\ (\pi_{id}, \cdots, \pi_{jd}\ ) = \hat{\delta}_0 + \hat{\delta}_1 \cdot \pi_{id} + \frac{e^{\hat{\delta}_2}}{1 + e^{\hat{\delta}_2}} \cdot \left(\frac{\sum_{j \neq i} \max\{\pi_{jd} - \pi_{id}, 0\}}{3} \right) \tag{4}$$

$$\hat{S}_{id}\ (\pi_{id}, \cdots, \pi_{jd}\ ) = \hat{\omega}_0 + \frac{1}{1 + e^{\hat{\omega}_1}} \cdot \pi_{id} + \frac{e^{\hat{\omega}_1}}{1 + e^{\hat{\omega}_1}} \cdot \sum_{j=1}^{4} \pi_{jd} \tag{5}$$

$$\hat{S}_{id}\ (\pi_{id}, \cdots, \pi_{jd}\ ) = \hat{\zeta}_0 + \frac{1}{1 + e^{\hat{\zeta}_1}} \cdot \pi_{id} + \frac{e^{\hat{\zeta}_1}}{1 + e^{\hat{\zeta}_1}} \cdot \min\{\pi_{1d}, \cdots, \pi_{4d}\} \tag{6}$$

In each model, we assume the satisfaction rating S_{id} is a measurement of the utility that subject i derives from distribution d. We further assume that, on top of the deterministic components of each utility function (viz., i's own material payoff and a given social motivation), there is an stochastic component following a normal distribution. This allows us to use nonlinear least squares to estimate the three models.

In the first model we impose $\hat{\delta}_1 = 1$. Notice that the function multiplying the inequality term in equation (4) is the logit function, which can only

13. Given the omission of α_i, all the distributions of income in the satisfaction ratings task involve situations with advantageous inequality.

take values in the range $[0, 1]$. Also, the functions multiplying one's own payoff and the social motivation in equations (5) and (6) ensure that each function takes a value between 0 and 1 and that the sum of the weights in each equation equals 1. This gives a natural interpretation to the estimated coefficients: $\frac{e^{\hat{\delta}_2}}{1+e^{\hat{\delta}_2}}$ is our structural estimate of β_i, $\frac{e^{\hat{\omega}_1}}{1+e^{\hat{\omega}_1}}$ is our structural estimate of ρ_i, and $\frac{e^{\hat{\zeta}_1}}{1+e^{\hat{\zeta}_1}}$ is our structural estimate of γ_i[14].

A crucial identifying assumption for estimating the parameters of social preferences is that satisfaction scores $\hat{S}_{id}$ will vary to some degree with the social motivations. Thus, we award a missing value to the social preference parameters for all subjects whose satisfaction scores do not vary with the social motivations[15].

B. Extrapolation to out-of-sample dictator games

A key goal of structural estimation is to document prevailing values of parameter estimates that give the best fit of a theoretical model to subject's observed data. Estimating different structural models can allow one two compare them, and state which fares best. Another important contribution of structural estimation is that it enables scientists to make predictions of a subject's behaviour in a set of choices of interest. This has become a primary tool for experimental economics in order to tackle criticisms of external validity. That is, as one makes experimental subjects face counterfactual, but different, situations of interest, one can use the elicited parameter values to make predictions based on optimal play, and analyse whether the fitted model has a high degree of out-of-sample success.

Social Preferences. Building on those strengths, we use the elicited parameter values of loss aversion and social preferences to make individual-level predictions for different models in the 22 binary dictator games. We couple loss aversion with the three social preference models presented earlier by (i) multiplying one's own material payoff by λ_i when in losses; and (ii) leaving their normal formulation when in gains. Denoting A_i as the set of strategies in a given dictator game, with $a_i \in A_i$ as its typical element, we compute, for every combination of subject, theory, and dictator game, the strategy a_i^* that maximises each utility function given the individual-level elicited parameters[16].

Moral Rules. Additionally, we use the data from the moral judgments tasks to make predictions for every combination of subject, theory, and

14. As monotonic transformations of utility functions do not alter the ranking of strategies, we opt not to restrict the constants in equations (4), (5), and (6) to taking the value of 0. This ensures we do not run into the statistical problems of regression through the origin whilst getting accurate estimates of the parameters of interest.

15. This affects 27 subjects in the lab experiment and 51 subjects in the online experiment. Around 12% of the subjects do not vary their satisfaction in all the scenarios, a percentage in line with the results reported in Bruhin, Fehr, and Schunk (2018).

16. In the appendix, we provide several propositions showing the best responses in the dictator games over gains and losses. We additionally demonstrate that, in the absence of loss aversion, the utility-maximizing action of social preferences is the same over gains and losses.

dictator game, of two moral rules, blame avoidance and praise seeking, developed in Gavassa-Pérez (2022). According to this theoretical framework, subjects will decide not between all possible alternatives but only between those that are deemed as morally appropriate from an impartial perspective. Thus, in effect moral judgments are used to reveal moral constraints on the strategy space of the dictator. The moral rule of blame avoidance (resp. praise seeking) states that anything with a negative moral rating (resp. only the alternatives with the highest moral rating) should not be taken into account in the decision-making process. Subject to that, an agent maximises their own utility, which is assumed to be that of a classical, self-regarding maximiser.

Letting R_i be the set of strategies compatible with a generic moral rule R for subject i, the main difference relative to traditional models of social preferences is that one computes the optimal strategy from the set $R_i \subseteq S_i$ given a utility function U_i. This opens the door to the existence of counter-preferential choices (relative to U_i) as far as the optimal strategy according to utility U_i is not an element of R_i. This theoretical framework can capture, in a simple, tractable structure, the spirit of the idea of *ethical preferences* in Harsanyi (1955), the concept of *commitment* in Sen (1977) , or the *Non-MaxU* framework presented in Smith and Wilson (2019).

C. Statistical Analysis

We relegate the discussion of the descriptive statistics for the results section, and focus here on summarising the three main statistical techniques we use to analyse the data of the modified dictator games.

1. Regression Modelling

We run regressions based on the following panel data, random effects logit model:

$$Pr\left[a_{it} = egal \mid \theta_i, b_t, \boldsymbol{P}_{it}^{'}, \boldsymbol{O}_{it}^{'}, \boldsymbol{S}_{it}^{'}\right] = \Lambda\left(\beta_0 + \beta_1 \cdot \mathbb{1}_{losses} + \beta_2 \cdot b_t + \beta_3 \cdot \mathbb{1}_{losses} \cdot b_t + \boldsymbol{P}_{it}^{'} \cdot \boldsymbol{\beta}_4 + \boldsymbol{O}_{i}^{'} \cdot \boldsymbol{\beta}_5 + \boldsymbol{S}_{i}^{'} \cdot \boldsymbol{\beta}_6 + \theta_i + \varepsilon_{it}\right) \quad (7)$$

The dependent variable is a dummy that takes the value 1 for person i in a dictator game t when that person chooses the egalitarian option in game t ($a_{it} = egal$), and 0 otherwise. The baseline model only includes as independent variables a dummy taking the value 1 when the game t involves redistributing losses ($\mathbb{1}_{losses}$), a variable (b) that captures the increase (relative to the unequal distribution) in the passive agent's payoff if the egalitarian distribution is chosen in game t, and an interaction between both. We extend the baseline model by including sequentially three vectors: (i) the out-of-sample predictions for each subject of each of the theories considered (henceforth, vector $\boldsymbol{P}_{it}^{'}$); (ii) dummies to control for the order effects in

which the tasks were presented (henceforth, vector $\boldsymbol{O}_i^{'}$); and (iii) sociodemographic controls (henceforth, vector $\boldsymbol{S}_i^{'}$). Additionally, we include a variable capturing the difference between the moral judgment of choosing the egalitarian and the unequal option $(mj(a_{it} = egal) - mj(a_{it} = uneq))$ in the vector $\boldsymbol{P}_{it}^{'}$ to explore how the intensity in moral judgments influences behavior. We present the average marginal effects of all models in the main text[17].

2. Theories Against the Void

A straightforward way of analysing our data is comparing each theory's degree of success in extrapolating behavior to the 22 dictator games. As a first step, we test each of the theories' success against a random benchmark. Letting one trial (in our case, one dictator game) follow a bernoulli distribution, with probability of success p, it follows that the sum of k successes of the $n = 22$ trials follows a binomial distribution with probability mass function defined by

$$Pr\left(k, n, p\right) = \frac{n!}{k!\left(n-k\right)!} \cdot p^k \cdot \left(1-p\right)^{n-k} \tag{8}$$

For our randomness benchmark, we use the natural assumption $p = 0.5$, since the dictator games are binary. We compare the distribution of the sum of successes for the 305 (resp. 348) subjects in the Lab (resp. Representative Sample) experiment against the randomness benchmark of a binomial distribution with parameters $n = 22$ and $p = 0.5$. In the next section, we report the outcome of the Kolmogorov-Smirnov one-sample test for each of the theories, alongside the histogram of the sum of successes of each theory plotted alongside the randomness benchmark for reference.

As a more stringent test, we use several binomial tests to compare the empirical proportions of 22 successes (i.e., 100% accuracy) against the theoretical proportion from the randonmess benchmark discussed above (i.e., $Pr\left(k = 22, n = 22, p = 0.5\right)$). Additionally, we use pairwise McNemar's tests and one Cochran's Q test to compare the empirical proportions of 22 successes of all theories against each other, as those tests control for the dependency of predictions between theories. This set of tests allows us to tell whether all theories are equally likely to have 100% accuracy in the extrapolation exercise.

3. Theories Against each other

The previous analysis presumes there is only one alternative theory to randomness. Thus, all successes of a theory are unambiguously credited to that theory. Yet, in reality there might be other theories that can predict

17. For robustness, we report the same analysis with population averaged and fixed effects specifications of the logistic panel data models in the online appendix.

the same data. If so, the previous tests are too lenient as they accept as evidence in favour of a theory any success, regardless of whether it could also be accounted for by an alternative theory. To account for this, we structurally estimate the probability of extrapolation success of each theory via a dirichlet-multinomial distribution, which we introduce below.

The Multinomial Part. We start assuming, as before, that each dictator game choice is a trial. This time, however, we assume different theories can potentially succeed in predicting the outcome of the dictator game. Let us denote N as the trials, T (with typical element $t \in T$) as the set of all candidate theories (i.e., the social preferences, moral rules, and a benchmark for the selfish homo economicus), k_t as the successes of theory t, and p_t as the parameter denoting the probability of success per trial of theory t. We model the probability of observing the vector of successes $\boldsymbol{k} = \{k_1, \cdots, k_t\}$ in N trials, given the corresponding vector of parameters $\boldsymbol{p} = \{p_1, \cdots, p_t\}$, as a multinomial distribution with density function given by

$$Pr\left(k_1, \cdots, k_t; N; p_1, \cdots, p_t\right) = \frac{N!}{\Pi_{t=1}^{T} k_t!} \cdot \Pi_{t=1}^{T} p_t^{k_t} \tag{9}$$

Where $\sum_{t=1}^{T} k_t = N$ and $\sum_{t=1}^{T} p_t = 1$ are identifying assumptions, implying that only one of the theories within T can succeed in explaining a given dictator game choice. This implication is unappealing for our setting, given that we observe some empirical correlation between the predictions of all theories.

The Dirichlet Part. In order to control for the over-dispersion in the data that is in contradiction with the last two assumptions, we modify the previous distribution in one important way. That is, we relax the assumption of parameters p_t being fixed, and we model them as being generated from a Dirichlet distribution, with hyper-parameters given by the vector of prior probabilities $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_t\}$. The probability of observing $\boldsymbol{p}$ given $\boldsymbol{\alpha}$ is now assumed to vary from trial to trial, and is given by the density function

$$Pr\left(p_1, \cdots, p_t; \alpha_1, \cdots, \alpha_t\right) = \frac{\Gamma\left(\sum_{t=1}^{T} \alpha_t\right)}{\Pi_{t=1}^{T} \Gamma\left(\alpha_t\right)} \cdot \Pi_{t=1}^{T} p_t^{\alpha_t - 1} \tag{10}$$

The Dirichlet-Multinomial Distribution. We can give an expression for the likelihood of the vector $\boldsymbol{k}$ in terms of p_t by multiplying both distributions, integrating over $\boldsymbol{p}$, and algebraically manipulating the resulting expression to get

$$\mathcal{L}\left(\boldsymbol{k}; \boldsymbol{p}; \rho\right) = \frac{\Pi_{t=1}^{T} \Pi_{r=1}^{k_t} \left(p_t \cdot (1-\rho) + (r-1) \cdot \rho\right)}{\Pi_{r=1}^{N} \left((1-\rho) + (r-1) \cdot \rho\right)} \tag{11}$$

Where $\rho = \frac{1}{1+\sum_{t=1}^{T}\alpha_t}$ is a parameter capturing the over-dispersion of successes. We use Maximum Likelihood and method of moments over the previous function to retrieve two estimates of $\hat{\boldsymbol{p}}$ and $\hat{\rho}$[18].

IV. Results

A. *Descriptive Results*

1. Calibrated Parameters

Fig. 1 reports Violin Plots of estimated parameters of social preferences and loss aversion for both the Lab and the Representative Sample experiments[19]. Dashed lines highlight the theoretical boundaries of social preferences parameters. The red, dashed line also serves as a benchmark for no disparity between *wta* and *wtp* (i.e., $\lambda_i = 1$).

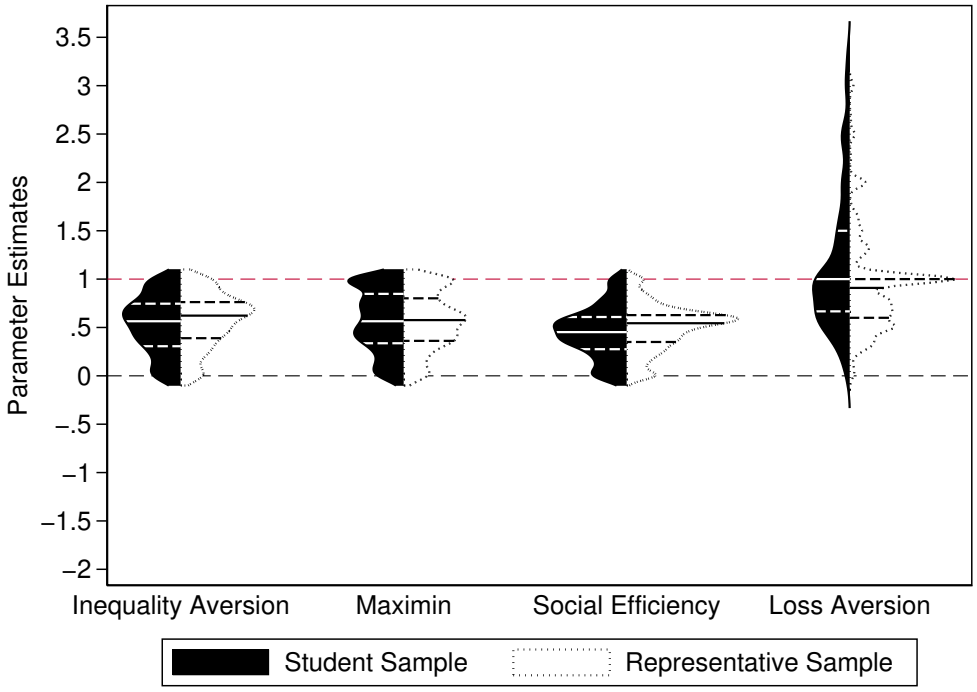


FIG. 1.—Violin Plots of the calibrated parameters. *We use equation* (4) *to measure the inequality aversion parameter* β_i *of equation* (1)*, equation* (5) *to measure the social efficiency parameter* ρ_i *of equation* (2)*, and equation* (6) *to measure the maximin parameter* γ_i *of equation* (3)*. Additionally, we use* $\hat{\lambda}_i = \frac{wta}{wtp}$ *as our measurement of the parameter* λ_i *of loss aversion. All parameters are measured at the individual level, and the violin plots represent the individual variation of each parameter in the student (black area) and the representative (white area) samples.*

18. For the initial paper on the Dirichlet-Multinomial distribution, see Mosimann (1962). For more details on the mathematics behind the dirichlet-multinomial distribution, and the different algorithms used for its maximum likelihood estimation, see Yu and Shaw (2014). We follow Weir and Hill (2002) in our estimation procedures.

19. We only use data where $\hat{\lambda}_i \leq 3.5$ for rendering the violin plot of loss aversion to get a more compact visual representation of the graph. For the statistical tests comparing the distributions of loss aversion, we use all the available data.

Inequality Aversion parameters have been extensively estimated (see Nunnari and Pozzi (2022) for a meta-analysis). In our sample, we observe a fairly important amount of people with a high level of advantageous inequality, replicating previous findings in the literature. We use several χ^2 tests and Bonferroni-corrected p-values (reported as p_b onwards) to compare the elicited distributions of β_i in our experiments against distributions reported in three relevant papers. Overall, we do not find significant differences between our elicited distributions and those reported in the literature when the subject pool over which the elicitation occurs is most similar[20]. We also find no significant differences between the elicited distributions of both experiments ($\chi^2 = 3.11$; $p_b = 1$). Taken together, the observed data strongly suggests that the self-reported elicitation method we used did not influence the distribution of social preference parameters.

The distributions of Maximin and Social Efficiency parameters are fairly similar across experiments. We observe that a substantial amount of people puts more than half the weight of their utility on each social concern, replicating the high degree of prosocial tendency that has been documented several times in the literature.

Finally, looking at the distributions of loss aversion we observe approximately half the subjects with loss aversion (i.e., $\lambda_i \geq 1$) in both experiments. Our elicited distributions of the loss aversion parameter are significantly different from the ones reported in Gächter, Johnson, and Herrmann (2022)[21], and are also significantly different from each other ($\chi^2 = 10.16$; $p_b = 0.00$). However, our individual-level means (Lab: 13.37; Representative Sample: 9.55) lie within the estimate boundaries for the loss aversion coefficient provided in Brown et al. (2024) for the UK.

2. Moral Judgments

Figs. 2 and 3 present the data elicited with the moral judgments tasks. Figs. 2a and 2b display the average moral ratings (vertical axis) our subjects gave to Person A in the 22 dictator games as a function of the efficiency of the dictator game (i.e., $\frac{b}{a}$. Horizontal axis). Each panel separates the average ratings according to the action done by Person A into two subpanels (left subpanels: Person A chooses the unequal distribution; right subpanels: Person A chooses the egalitarian distribution). Furthermore, we plot the data of the gains and losses domains with two different lines (black cir-

20. More specifically, we do not find significant differences between the distribution in the Lab experiment and the distributions elicited in the lab experiments reported in Blanco et al. (2011. $\chi^2 = 3.47$; $p_b = 1$) and Beranek et al. (2017. Nottingham Students: $\chi^2 = 0.89$; $p_b = 1$). We also do not find differences between the distribution elicited in the Representative Sample and the distribution elicited online by Beranek et al. (2017. $\chi^2 = 5.21$; $p_b = 0.74$). However, both our elicited distributions differ from the distribution proposed in Fehr and Schmidt (1999. Lab: $\chi^2 = 11.58$; $p_b = 0.03$; Representative Sample: $\chi^2 = 17.86$; $p_b = 0.00$) at the 5% level, and the distribution of the Representative Sample significantly differs from the one elicited in the lab by Blanco et al. (2011. $\chi^2 = 9.35$; $p_b = 0.09$) at the 10% significance level.

21. Lab: $\chi^2 = 66.80$; $p_b = 0.00$; Representative Sample: $\chi^2 = 127.64$; $p_b = 0.00$)

cles: gains; white circles: losses). Panel Fig. 2a displays the data of the Lab experiment and Fig. 2b displays the data of the Representative Sample experiment. Fig. 3 presents the average differences between the moral judgment of choosing the egalitarian and the unequal distribution (horizontal axis) as a function of the efficiency of the dictator game (horizontal axis). We plot these differences separately for the losses and the gains domain (black circles: gains; white circles: losses).

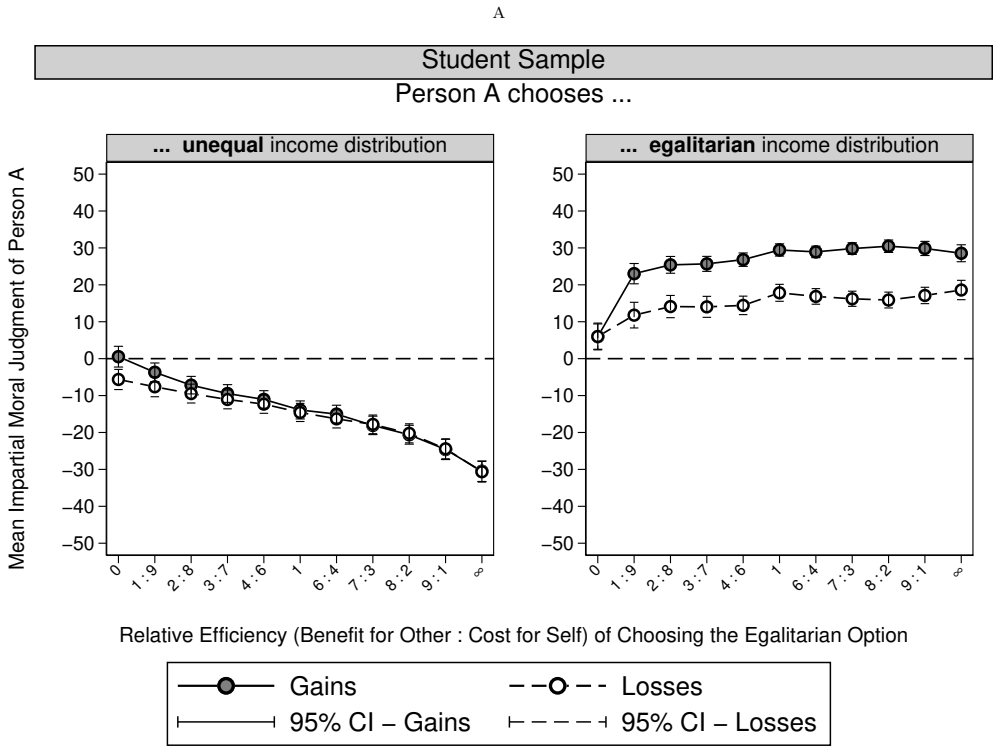
Figs. 2a and 2b highlight three main features of subjects' moral perceptions of dictator games. First, moral judgments are strongly different from zero in most cases, highlighting that people see distributional choices as of moral importance. Second, the negative (resp. positive) moral ratings of choosing the unequal (resp. egalitarian) distribution suggests that both the positive and negative domain of morality are important to subject's understanding of dictator games: being selfish entails moral condemnation meanwhile giving is judged as worthy of praise. Third, we document an important asymmetry on the role of the gains-losses distinction in the morality of dictator games. Meanwhile choosing the unequal distribution is seen as equally blameworthy under gains and losses, choosing the egalitarian distribution is perceived as more praiseworthy over the gains domain.

Fig. 3 provides two main additional findings. Eyeballing the figure, one can observe that the distance between the moral perceptions of choosing unequal and egalitarian distributions is increasing in the efficiency of the game. This result is important as it implies (i) that there is a clearer distinction in the moral domain between both alternative choices as efficiency increases; and (ii) that the intensity of the moral preference of the egalitarian distribution is increasing in the efficiency of a dictator game. Furthermore, this effect is mediated by the domain of play: dictator games over gains display a greater sensitivity to efficiency, thereby making the increase in moral distance between choosing the unequal and the egalitarian option greater than it is over losses as efficiency increases. What is most striking is that the previous five findings scale from the Lab experiment to the Representative Sample, providing robust evidence of the landscape of moral perception of distributional situations[22].

3. Dictator Game Play

Fig. 4 presents a descriptive summary of the dictator games choices. More specifically, we focus on the switching points (i.e., the point where a subject switches from choosing the unequal to choosing the egalitarian distribution) in the gains and losses domains. Fig. 4a presents data from the Lab experiment and Fig. 4b presents data from the Representative Sample. Both figures are divided into two subpanels. The left subpanels present the empirical cumulative frequency (vertical axis) of the switching point as a function

22. All five findings are backed by pre-registered statistical tests that we provide in the online materials. All p-values presented are Bonferroni-corrected.



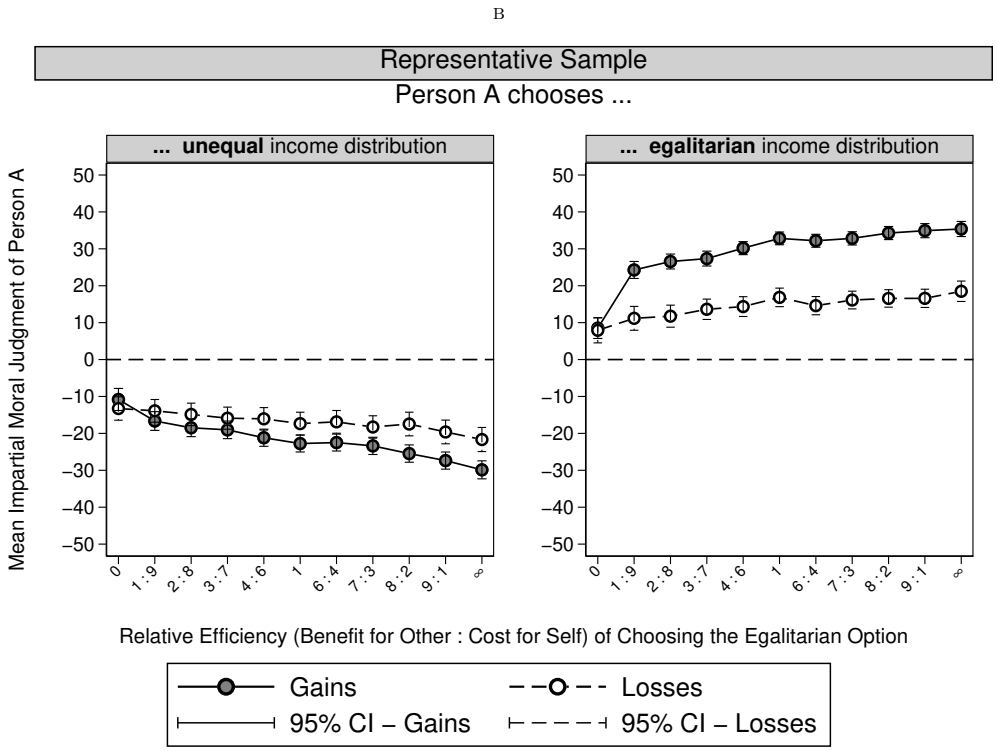


FIG. 2.—Average Moral Judgments of Dictator Games

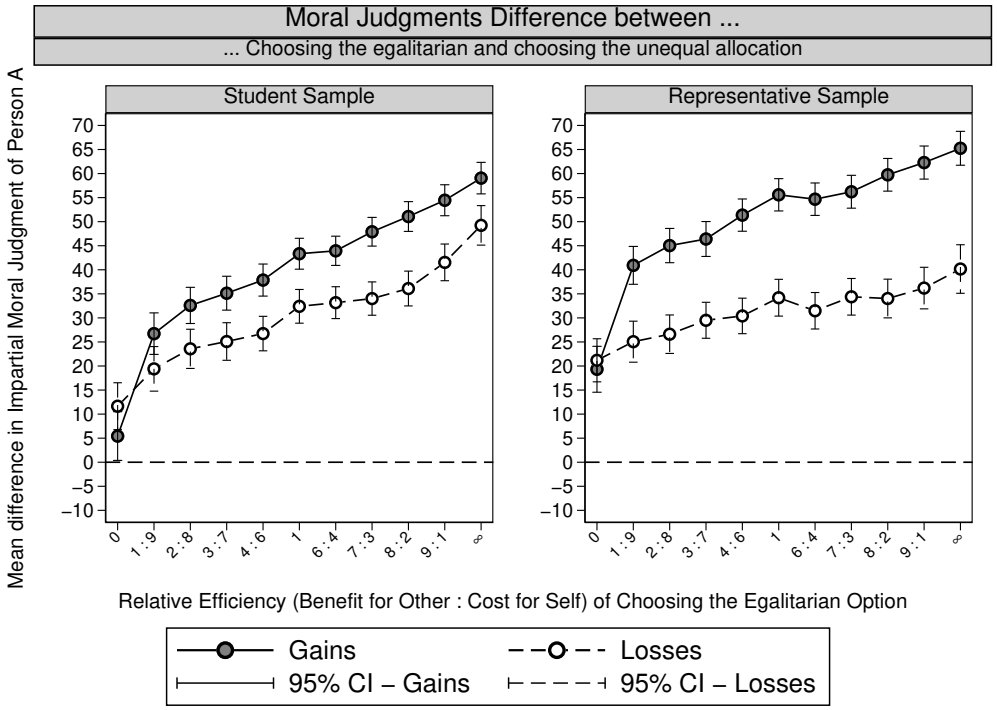


FIG. 3.—Average difference between the moral judgments of the egalitarian and the unequal distribution

of the efficiency of the dictator games (horizontal axis). Lines with black (resp. white) circles represent the gains (resp. losses) domain data. The right subpanels exploit the within-subject nature of the data to present, in a heatmap, the joint distribution of efficiency levels at which switching points occur in the gains (vertical axis) and losses (horizontal axis) domain. We plot a 45-degree line for reference, representing equal switching points over gains and losses. Disregarding loss aversion, social preference models can only predict data along this line. Data below the 45-degree line comes from subjects being more selfish over losses than over gains (i.e., where switching to the egalitarian distribution happens at lower efficiency levels in the gains domain), and hence is, prima facie, compatible with social preferences coupled with loss aversion.

Based on the results of signed-rank tests, we can conclude that the distribution of switching points over gains and losses are significantly different in both experiments (Lab: $Z = -443.91$; $p_b = 0.00$; Representative Sample: $Z = -683.71$; $p_b = 0.00$). Switching to the egalitarian distribution tends to occur earlier in the efficiency scale over gains. This is accentuated in the Representative Sample, where we find a greater frequency of data below the 45-degree line in the heatplot of the Representative Sample relative to that of the Lab data. This is counter to what we would expect if loss aversion and social preferences were the driver of choices in the dictator games; as, given the distribution of the loss aversion parameter reported in Fig. 1, loss

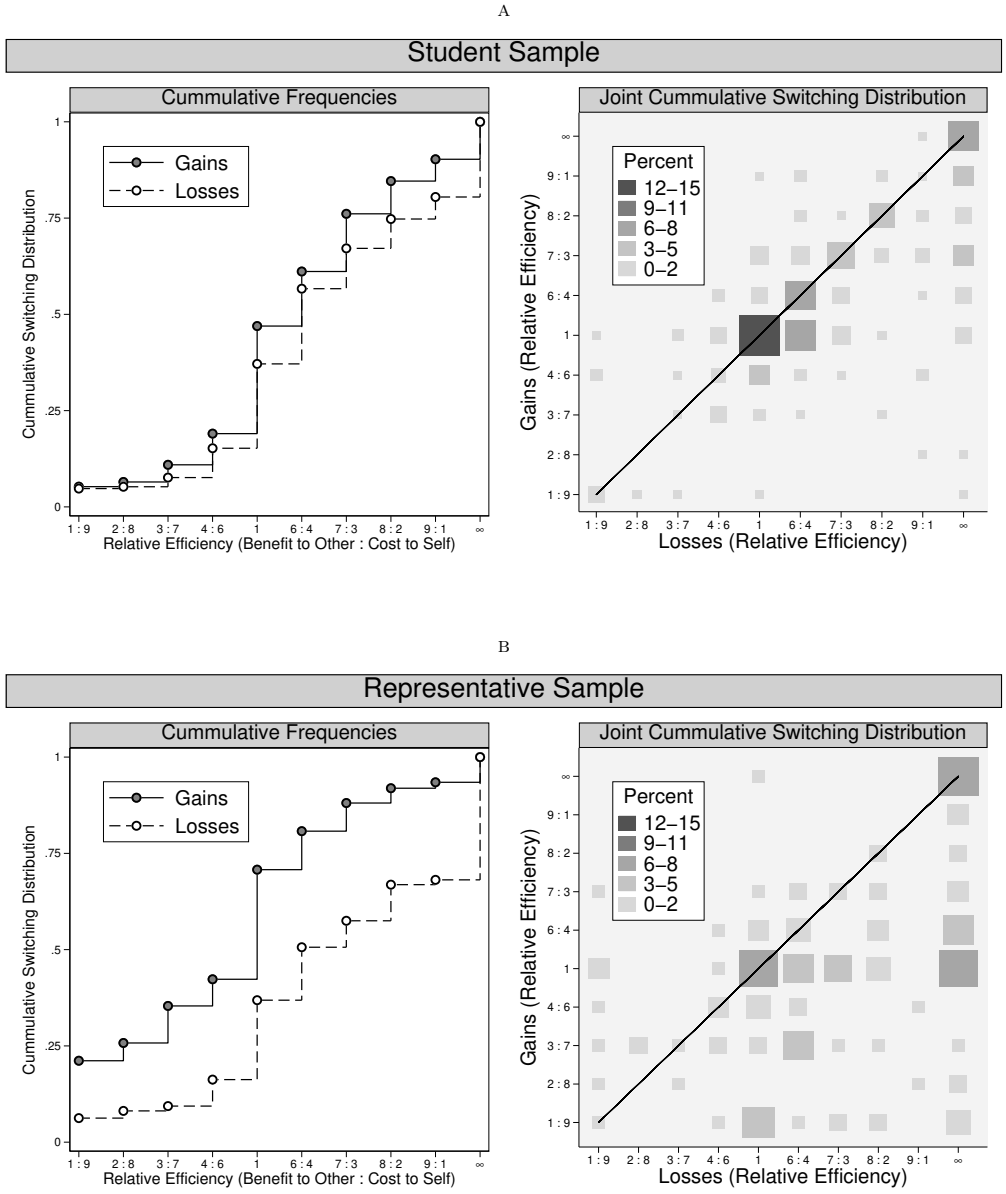


FIG. 4.—Cummulative frequencies and heatplots of switching points over gains and losses Dictator Games

aversion is more frequent in the Lab than in the Representative Sample data.

B. Regression Estimates

Table 2 reports the coefficients and average marginal effects (integrating out the error term θ_i) of the panel data, logistic regressions presented above. This allows us to study the determinants of the probability of choosing the egalitarian distribution. We run four specifications for both the Lab and the Representative Sample experimental data. Each specification is defined by the last four rows in the table (i.e., error specification and variables included). We accompany the table with Figs. 5 and 6. Both figures report the predicted probabilities of models in column (4) and (4′) evaluated at different levels. In the case of Fig. 5, we evaluate the predicted probabilities (vertical axis) of the Lab (left subpanel) and the Representative Sample (right subpanel) data as a function of b (horizontal axis) and $\mathbb{1}_{losses}$ (gains: dark circles; losses: white circles). In the case of Fig. 6, we evaluate the predicted probabilities (vertical axis) of the Lab (top row) and the Representative Sample (bottom row) data as a function of (i) each theory's prediction (solid line: theory predicts unequal distribution should be chosen; dashed line: theory predicts egalitarian distribution should be chosen); (ii) b (horizontal axis); and (iii) $\mathbb{1}_{losses}$ (gains: left subpanel; losses: right subpanel).

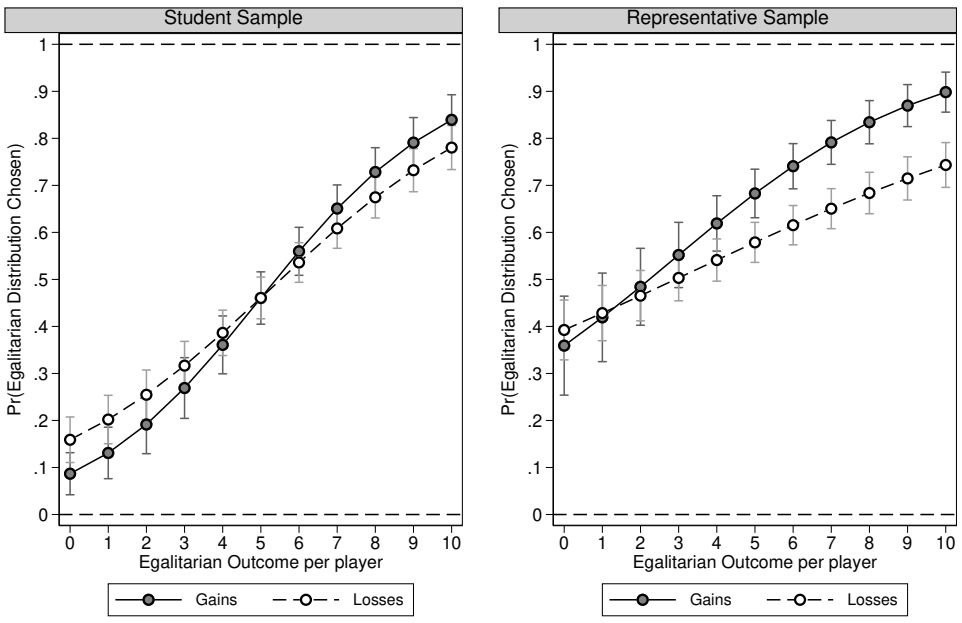


FIG. 5.—Predicted probabilities evaluated at every value of b and $\mathbb{1}_{losses}$

All the variables related to the game features (viz., b, $\mathbb{1}_{losses}$, and their interaction) have significant coefficients at the 1% level in model (1). This effect is robust to the inclusion of the theoretical predictions of each theory,

Table 2
Panel Data estimates and Average Marginal Effects of the logistic regressions

	Student Sample		Representative Sample		Student Sample		Representative Sample		Student Sample		Representative Sample		Student Sample		Representative Sample	
	(1)		(1')		(2)		(2')		(3)		(3')		(4)		(4')	
	β / SE	AME	β / SE	AME	β / SE	AME	β / SE	AME	β / SE	AME	β / SE	AME	β / SE	AME	β / SE	AME
Game Features																
b	0.805***	0.072***	0.660***	0.049***	0.691***	0.062***	0.561***	0.041***	0.697***	0.062***	0.562***	0.041***	0.306***	0.064***	0.532***	0.043***
	(0.051)		(0.040)		(0.056)		(0.049)		(0.058)		(0.049)		(0.113)		(0.124)	
losses	1.563***	−0.001	1.513***	−0.069***	1.392***	0.002	1.424***	−0.047**	1.354***	0.000	1.428***	−0.047**	2.101**	0.010	2.800***	−0.068***
	(0.274)		(0.254)		(0.291)		(0.292)		(0.304)		(0.292)		(0.703)		(0.614)	
$b \times losses$	-0.312***		−0.454***		−0.273***		−0.396***		−0.267***		−0.397***		−0.238*		−0.417***	
	(0.044)		(0.043)		(0.048)		(0.047)		(0.049)		(0.047)		(0.141)		(0.137)	
Moral Rules																
Blame Avoidance					0.680***	0.079***	−0.954***	−0.111***	0.655***	0.076***	−0.960***	−0.112***	0.582	0.072***	−0.301	−0.108***
					(0.208)		(0.302)		(0.217)		(0.303)		(0.549)		(0.539)	
Praise Seeking					−0.503*	−0.057*	0.000	0.000	−0.377	−0.042	0.012	0.001	−1.067	0.006	0.430	0.043
					(0.305)		(0.305)		(0.319)		(0.305)		(0.730)		(0.656)	
$mj(egal) - mj(uneq)$					0.008*	0.001*	0.020***	0.002***	0.007	0.001*	0.020***	0.002***	0.012	0.001*	0.020***	0.002***
					(0.004)		(0.003)		(0.004)		(0.003)		(0.009)		(0.007)	
Social Preferences																
Inequality Aversion					0.241	0.028	0.233	0.029	0.277	0.032	0.228	0.028	−0.964	0.022	−0.528	0.041*
					(0.188)		(0.164)		(0.192)		(0.164)		(0.608)		(0.560)	
Maximin					0.145	0.017	0.366*	0.045	0.094	0.011	0.361	0.045	−0.683	0.029	0.130	0.031
					(0.214)		(0.221)		(0.224)		(0.221)		(0.527)		(0.584)	
Social Efficiency					0.346**	0.041**	0.100	0.012	0.349**	0.041**	0.100	0.012	1.240	0.008	2.617***	0.047**
					(0.160)		(0.126)		(0.167)		(0.126)		(0.839)		(0.752)	
Constant	-4.073***		−2.098***		−4.181***		−2.257***		−4.546***		−2.596***		−3.562***		−3.329***	
	(0.308)		(0.206)		(0.371)		(0.248)		(0.470)		(0.314)		(0.574)		(0.526)	
Clusters	305		348		305		348		288		348		288		348	
Panel Data Model	RE		RE		RE		RE		RE		RE		RE		RE	
Predictions	✗		✗		✓		✓		✓		✓		✓		✓	
Order Dummies	✗		✗		✗		✗		✓		✓		✓		✓	
3−Way Interactions	✗		✗		✗		✗		✗		✗		✓		✓	

Note.—The table reports the regression coefficients, the average marginal effects, and the standard errors (in parentheses) of several logit, random effects regression models. The dependent variable is a dummy that takes the value 1 when a subject chooses the egalitarian distribution and 0 otherwise. In models (1) and (1′), we include as independent variables b, which is defined as the gain of the receiver (relative to the unequal distribution outcomes) when the egalitarian distribution is chosen; *losses*, which is a dummy that takes the value 1 when the binary dictator game involves losses and 0 otherwise; and an interaction between both. In models (2) and (2′), we include the theoretical predictions of social preference and moral rules models as independent variables . In models (3) and (3′), we include dummies to control for order effects. In models (4) and (4′), we include three-way interactions between b, *losses*, and each of the theoretical predictions. We omit order effects and three-way interactions from the output to make the table more compact. The bottom of the table acknowledges the number of subjects included in each regression. * $p < .1$, ** $p < .05$, *** $p < .01$.

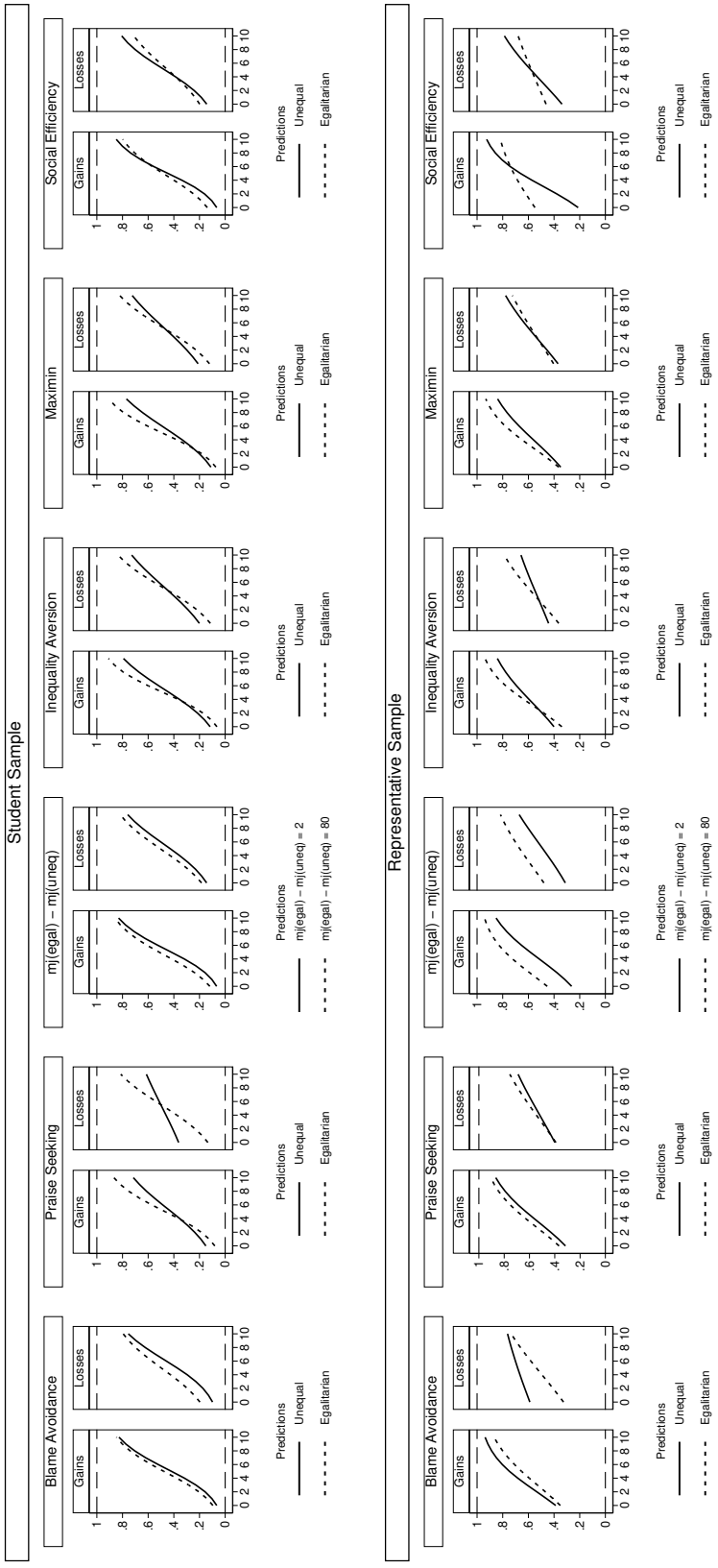


FIG. 6.—Predicted Probabilities of each theory

and order effects; and scales to the Representative Sample data. When we include three-way interactions, the coefficient of $b \times \mathbb{1}_{losses}$ is still significant for the Lab data at the 10% level, and remains significant at the 1% level for the Representative Sample. The average marginal effect of b is always positive, and ranges between 6.2 and 7.3% (resp. 4.1 and 4.9%) for the Lab (resp. Representative Sample), implying that, ceteris paribus, an increase of 1 unit in the payoff of the egalitarian distribution increases the probability of choosing the egalitarian distribution by the aforementioned percentages. The average marginal effect of $\mathbb{1}_{losses}$ is not significant in the Lab, but significant in the Representative Sample data; and it ranges between -0.1 and 1% (resp. -4.7 and -6.9%) in the Lab (resp. Representative Sample data).

Looking at Fig. 5, we can see that the predicted probabilities of choosing the egalitarian outcome are increasing in b, confirming the previous results. Also, in the Lab experiment predicted probabilities are higher for losses at lower levels of b, but higher for gains at higher levels of b. This gives a rationale for the lack of a significant effect of AME in the Lab data: positive and negative effects at different levels of b cancel out. Finally, in the Representative Sample data we see that the predicted probabilities of choosing the egalitarian outcome are always higher in the gains domain, supporting the finding of a positive and significant AME. In short, we find that an increase in b increases the probability of choosing the egalitarian outcome; and that there is a higher probability of choosing the egalitarian outcome when redistributions are made over gains, especially in the Representative Sample.

Moral Rules are important determinants of redistributive choices. The rule of Blame Avoidance has a positive and significant coefficient in model (1). This effect is robust to the inclusion of order effects. In terms of average marginal effects, when Blame Avoidance predicts the choice of an egalitarian distribution, the probability that the egalitarian distribution is chosen in the Lab data increases, on average, by 7.2 to 7.9%. Strangely, the effect of Blame Avoidance on egalitarian choices reverses in sign for the Representative Sample data. As we document in the appendix, this sign reversal disappears when we restrict our sample to individuals for which Blame Avoidance gets at least 50% of the dictator games choices right. The rule of Praise seeking performs worse than that of blame Avoidance, as it has a negative sign in all specifications for the Lab data; and its statistical significance, when present, is weak. In contrast, the intensity of the moral preference of the egalitarian over the unequal distribution is the most powerful determinant of egalitarian choices in dictator games. Its coefficient has a positive sign and statistical significance, it is robust to the inclusion of order effects, and its effect scales to the Representative Sample data. Whenever the intensity of moral preference for the egalitarian distribution increases by 100 points, the probability of choosing the egalitarian option increases, on average, by 10% (resp. 20%) in the Lab (resp. Representative Sample) data.

Social Preferences also influence redistributive choices. Table 2 shows the coefficient of Maximin is weakly significant in the Representative Sample, although its significance is not robust to the inclusion of order effects and three-way interactions. Social Efficiency is the only social preference with statistically significant average marginal effects at the 5% level, evidencing it is a determinant of egalitarian choices. Its significance is robust to the inclusion of order effects in the Lab data. Once we include three-way interactions, Social Efficiency becomes significant at the 5% level in the Representative Sample data, and becomes insignificant in the Lab specification. The average marginal effects of social preferences range between 1.1 and 4.7%. Overall, the effects of social preferences are less robust to the inclusion of order effects and three-way interactions than the effects of moral rules.

Fig. 6 provides additional insights on the effect of each theory on egalitarian choices, as using three-way interactions allows each theory to have a potentially different effect for each combination of b and $\mathbb{1}_{losses}$. In the Lab data, some theories, like Praise Seeking, Inequality Aversion, and Maximin have an effect on the probability of choosing the egalitarian distribution that is moderated by b, showing the importance of accounting for interaction effects. In the case of Inequality Aversion, this moderation is also present in the Representative Sample data.

C. Robustness Checks

1. Theories Against the Void

Fig. 7 reports the frequency (vertical axis) of the sum of successes of each theory (histogram with red outline) and compares it against a randomness benchmark (i.e., histogram with black outline. A binomial distribution with $k = 22$ and $p = 0.5$). Each panel presents the data for each theory independently; the top row reporting data from the Lab and the bottom row reporting data from the Representative Sample experiment.

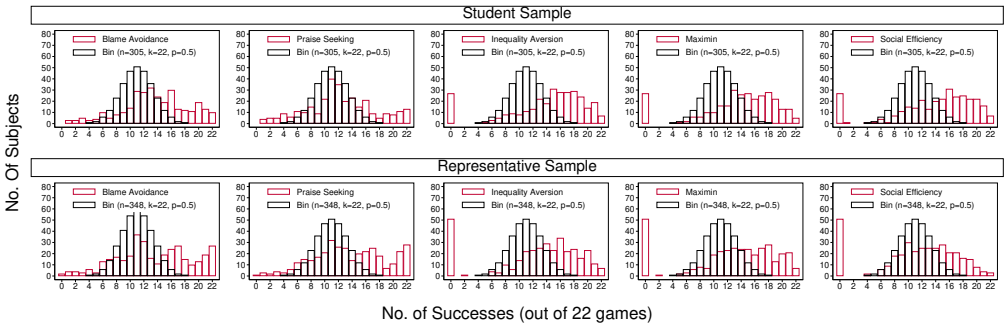


FIG. 7.—Histogram of the sum of successes of each theory

According to Kolmogorov-Smirnov one-sample tests, the distribution of

each theory is significantly different from the randomness benchmark[23]. Additionally, the empirical proportions of 22 successes (i.e., 100% consistency) for each theory is significantly different from the theoretical proportion of the randomness benchmark. Intuitively, this is confirmed by looking at how the histogram with red outline, in each panel, is more populated in the right tail (i.e., a higher sum of successes) relative to the histogram with black outlines. All these patterns show up in both the Lab and the Representative Sample data. Taken together, the Lab results – and its replication in a general population – suggest that all the theories we study have empirical content beyond what could be accounted for by randomness.

Given that all the theories fare better than randomness, a natural next step is to see whether they predict the same data, or whether they are complementary in explaining distributional choices. As an exploratory analysis, and to see whether all theories are doing well by correlation (viz., because they predict the data of the same subjects), we present an additional graph. Fig. 8 presents heatplots of the sum of successes of each theory (vertical axis) against each subject (horizontal axis) for the Lab (left panel) and the Representative Sample (right panel). We sort subjects by the sum of successes of Blame Avoidance. The graph is quite straightforward in revealing that Moral Rules and Social Preferences are complementary in our understanding of distributional choices: whereas Blame Avoidance and Praise Seeking have a high degree of correlation in their sum of successes at the individual level, they correlate very poorly with the sum of successes of Social Preferences. The reverse also holds true.

Additionally, we compare each theory's empirical proportions of 22 successes (i.e.,100% consistency) with a Cochran's Q test, and find that they are not equally likely to get 22 successes (Lab: $Q = 16.27$; $p_b = 0.01$. Representative Sample: $Q = 49.83$; $p_b = 0.00$). Pairwise McNemar's tests suggest this difference is mainly driven by the difference between Social Preferences (i.e., Inequality Aversion and Maximin) and selfishness for the Lab sample, and by the difference between Moral Rules (i.e., Blame Avoidance and Praise Seeking) and Social Preferences (i.e., Inequality Aversion, Maximin, and Social Efficiency) in the Representative Sample. Overall, this result provides a slight edge to Moral Rules over Social Preferences, as it implies they get a significantly higher proportion of full consistency than Social Preferences in the Representative Sample.

2. Structural Estimation of a Horse-Race

We end the results section reporting the structural estimates of the proportions of the dirichlet-multinomial distribution, $\boldsymbol{p}$, and of ρ, for both the

23. Lab Experiment: Blame Avoidance ($D_{305} = 0.24$; $p_b = 0.00$); Praise Seeking ($D_{305} = 0.17$; $p_b = 0.00$); Inequality Aversion ($D_{305} = 0.19$; $p_b = 0.00$); Maximin ($D_{305} = 0.17$; $p_b = 0.00$); Social Efficiency ($D_{305} = 0.21$; $p_b = 0.00$). Representative Sample Experiment: Blame Avoidance ($D_{348} = 0.20$; $p_b = 0.00$); Praise Seeking ($D_{348} = 0.22$; $p_b = 0.00$); Inequality Aversion ($D_{348} = 0.11$; $p_b = 0.00$); Maximin ($D_{348} = 0.13$; $p_b = 0.01$); Social Efficiency ($D_{348} = 0.13$; $p_b = 0.00$).

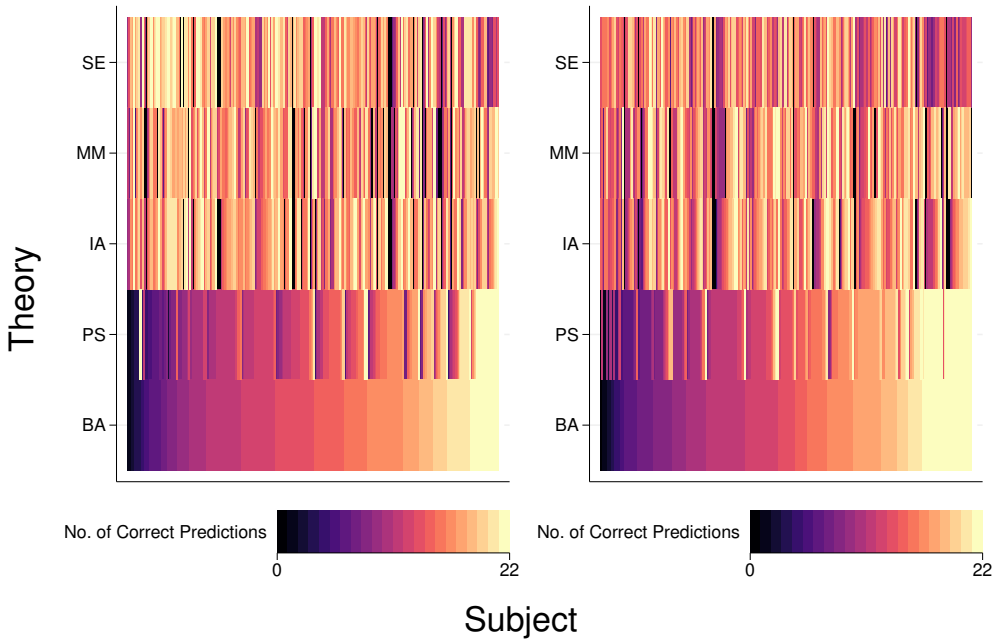


FIG. 8.—Heatplots of the sum of successes of each theory against each subject

Lab and the Representative Sample. Table 3 reports both Maximum Likelihood (MLE) and Method of Moments (MoM) estimates, alongside their respective standard errors.

The first thing worth noting is that, regardless of the estimation method, the proportions of Moral Rules and Social Preferences are higher than those of the Homo Economicus. This holds for both the Lab and the Representative Sample data. Second, the probabilities of Moral Rules and Social Preferences that we estimate from the two experiments are remarkably close, showing that the mechanisms of distributional choice scale from the lab to a general population. Third, the probabilities of Moral Rules are slightly higher than those of Social Preferences in the representative sample, highlighting the importance of impartial moral concerns in explaining the distributional choices of the general population. And fourth, the probability estimates of each individual theory are remarkably similar, suggesting that distributional choices are best understood as stemming from different motivations and rationales, rather than being driven by a unique, universal one. This implies that theories aiming to predict other-regarding behavior are closer to being complementaries, rather than substitutes, and that we need to take into account the heterogeneity in motivational forces to provide an accurate representation of why people redistribute.

TABLE 3

PARAMETER ESTIMATES OF THE DIRICHLET-MULTINOMIAL DISTRIBUTION

	Student Sample		Representative Sample	
	MLE	MoM	MLE	MoM
Moral Rules				
Blame Avoidance	0.176***	0.172**	0.195***	0.185*
	(0.004)	(0.082)	(0.005)	(0.097)
Praise Seeking	0.159***	0.162*	0.199***	0.187*
	(0.004)	(0.092)	(0.005)	(0.099)
Social Preferences				
Inequality Aversion	0.177***	0.179***	0.169***	0.175**
	(0.004)	(0.062)	(0.005)	(0.074)
Maximin	0.172***	0.174***	0.168***	0.174**
	(0.004)	(0.062)	(0.005)	(0.077)
Social Efficiency	0.152***	0.156**	0.169***	0.166**
	(0.004)	(0.066)	(0.005)	(0.072)
Selfishness				
Homo Economicus	0.142***	0.143	0.119***	0.123
	(0.004)	(0.096)	(0.004)	(0.129)
Overdispersion				
ρ	0.020***	0.016***	0.042***	0.027***
	(0.001)	(0.002)	(0.002)	(0.003)

NOTE.—The table reports the structural estimates of the proportions of each theory alongside the estimate of the parameter accounting for the over-dispersion of our data. Assuming we randomly draw the choice of a subject in one of the binary dictator games, the vector of proportions estimate the likelihood that each theory has of that subject following each theory. Standard errors are presented in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

V. Discussion and Concluding Remarks

We set out to experimentally investigate distributional behavior. The two main objectives of the paper are to provide evidence of whether the domain of play (i.e., gains and losses) influences choices, and document the mechanisms of distributional decisions. The experiment consists on presenting each subject 22 binary dictator games, half of them over gains and half of them over losses, and several other tasks aimed at structurally estimating parameters of several theories of social preferences and moral rules. We, then, use the elicited data in the latter set of tasks to extrapolate behavior to the binary dictator games.

The regression results show that game-specific features (viz., gain to others relative to the unequal distribution, domain of play) influence behavior. We find that the higher the benefit to the passive subject, the more likely it is that the egalitarian distribution in a dictator game is chosen. Addi-

tionally, we document an asymmetry in distributional behavior over gains and losses: subjects display more egalitarian behavior over gains and more selfish behavior over losses. Both effects persist even when controlling for the influence that a set of economic theories have on distributional choices.

Both the regression and the distributional analysis presented herein document the importance of social preferences and moral rules as drivers of distributional behavior. Each theory's degree of success in predicting redistribution outperforms what was expected by randomness, and is still important when taking into consideration all the alternative theories under investigation as potential drivers of distributional behavior. As opposed to what was assumed the norm a century ago in the economic discipline, our estimations suggest that the classical Homo Economicus accounts for 11 to 14% of the behavior in distributional choices. In contrast, we estimate that the alternative proposed in economics in the 90's and 2000's, that is, self-centered social preferences captured by concerns for inequality, maximin, and social efficiency, represent around 50% of our subjects. Finally, the disinterested moral framework presented in Gavassa-Pérez (2022) represents the behavior of the remaining 33 to 39% of our experimental subjects.

There are two outcomes of this paper that are worth discussing in some detail. First, our experimental results evidence the important degree of heterogeneity that underlies distributional behavior. This is in line with the findings of related papers, such as Andreoni and Miller (2002), Fisman, Kariv, and Markovits (2007), and Bruhin, Fehr, and Schunk (2018), who document different behavioral types rationalising prosocial behavior. However, those papers only consider preference-based behavior. We consider theories based on fundamentally different primitive assumptions, as the Moral Rules we consider build on the Spirit of the concept of *Ethical Preferences* in Harsanyi (1955), which is independent from one's utilty; the concept of*Commitment* in Sen (1977), which is counterpreferential in nature; and the *NonMaxU* theory advanced in Smith and Wilson (2017, 2019)[24]. Whilst we cannot rule out that the behavior captured by Moral Rules is violating axioms of preference, as the experiment is not designed to test for intransitivity, our experimental evidence and analysis documents (i) that there exist non-preferential paths to social behavior; (ii) that they can be captured in a simple, tractable framework; enabling accessible experimental tests of it; and (iii) that they can successfully represent the prosocial behavior of a non-trivial quantity of people; in most cases where the canonical models in the literature fail. This evidence, as the one reported in Gavassa-Pérez (2022), opens the door to a new strand of the literature that aims (i) to subsequently test whether this disinterested path to social doing can predict behavior of a broader set of games; and (ii) to build innovative

24. See, especially, Section IV's discussion in Harsanyi (1955, pages 314-316) for the distinction between Ethical Preferences, based on Social Welfare, and Utility. Sen (1977) is also very clear when he states that '*Commitment does involve ... counterpreferential choice, destroying the crucial assumption that a chosen alternative must be better than (or at least as good as) the others for the person choosing it*'. Finally, see chapter 3 in Smith and Wilson (2019) for a discussion of social preferences and behavior.

experiments (e.g., perhaps with the use of neuroscientific evidence) that can separate between both the preference-based and the non-preference-based accounts of altruistic behavior.

Second, perhaps the most surprising result is that the intensity of the moral preference of the egalitarian over the unequal distribution is a very strong determinant of the egalitarian choice. The concept of moral difference between the two options as a driver of choice has theoretical support from moral philosophy and economics, which we discuss below.

In moral philosophy, 'Characteristics' of Shaftesbury (2000) is traditionally seen as the starting point of moral sentimentalism, a philosophical school followed by David Hume and Adam Smith. One of the features that made Shaftesbury's take on ethics so peculiar was the so-called concept of 'second-order' moral judgments[25]. Shaftesbury defines the *Moral Sense* as that sense which produces like or dislike for our first-order sentiments. In that way, it is only because we judge Benevolence as likeable why it becomes a virtue. One can build up on this concept of second-order moral judgments to rationalise the intensity of moral preferences. Let's consider, for instance, someone facing a choice between two alternatives. He judges one alternative as mildly bad, and another one as mildly good. We now have to judge that person, knowing that he chose the mildly bad alternative. As the moral difference between those two actions was not great, it would not be natural to consider him a *villain*, or someone deeply vicious. If we now judge another person, who judged one alternative as deeply morally wrong, and another one as deeply morally good, yet he chose the former, it would be more natural to judge him as a *bad* person. Thus, it seems that the moral distance between alternatives can enter into second-order considerations, providing, thus, a rationale for action: if I want to avoid being a bad person, and it is more likely I'll be a bad person if I choose the morally worse action when the moral distance between alternatives increases, then the likelihood of me choosing the morally right action will be increasing in the moral distance between alternatives.

Within neoclassical economics, there has been a debate as to whether utility was cardinal or ordinal. Without entering into that debate, we want to highlight the work of Fishburn (1970) on *preference differences*, or preference *intensity*. Fishburn proposes a binary relation $\prec^*$ on pairs within the cartesian product of an action space, $A \times A$. If, let's say, we have elements $x, y, z \in A$, then we can model the concept of preference intensity with utility formulations $u(.)$ that follow either $(x, y) \prec^* (x, z) \Rightarrow u(x) - u(y) < u(x) - u(z)$ or $(x, y) \prec^* (x, z) \Leftrightarrow u(x) - u(y) < u(x) - u(z)$. This theoretical framework allows for preference intensity to convey some valuable information in a subject's utility. This is a fruitful framework to capture

25. This is explicit in the following passage: '*the affections of Pity, Kindness, ... being brought to mind by reflection, become Objects. So that, by means of this reflected sense, there arises another kind of Affection towards those Very Affections themselves, which have been already felt, and are now become the Subject of a new Liking or Dislike*' (*Characteristics, Vol.2; Sect. 3; [28]*).

the concept of intensity in the moral domain. By allowing for intensity in moral judgments, which we will call $\prec^*_{mj}$, to convey an important information about the appropriateness of said action, one can build a moral rule dependent on it, providing a theoretical basis for the link between moral intensity and egalitarian choices that we present in this paper. Furthermore, the moral difference we found is also closely related to the moral cost function discussed in List (2007), and the moral framework presented in Levitt and List (2007). One can interpret the results as providing empirical evidence for the existence of a morality dependent on the stake of external benefit to others (captured by v in Levitt and List (2007)), and for how such moral perception influences behavior in distributional choices.

Finally, the results we present herein are also compatible with a drift diffusion model of decision making where impartial moral concerns are an important factor influencing choices. In drift diffusion models as the one presented in Ratcliff and McKoon (2008), there are several important constructs: the real decision value, used as the measurement that assesses all the relevant alternatives; the reaction time, measured as the time taken between the initial stimuli and the decision; and the decision boundaries, defined as the upper and lower thresholds of a decision value to be reached in order for a subject to make a decision. These models have been used to capture a neurological representation of altruistic decision making, as in Hutcherson, Bushong, and Rangel (2015), where the real decision value is modelled as a weighted linear combination of the payoffs for oneself and for others. The fact that the moral distance between both choice alternatives influences the likelihood of choosing the egalitarian option, even when controlling for the predictions of other-regarding preferences, suggests that impartial moral perceptions can be an important neurological driver of distributional decisions.

Appendix

In this appendix we formally define the binary dictator games that are the core of our investigation. We, then, provide some propositions on the relation between parameter values of other-regarding preferences and play in the dictator games over gains and losses. Furthermore, we provide some corollaries to show that, in the absence of reference dependence, none of the other-regarding preference models considered can predict asymmetric choices under gains and losses for the same set of payoffs.

Preliminaries

The Binary Dictator Games. Let us define $\mathcal{I} := \{d, r\}$ as the **set of players**, using element d to refer to a dictator and element r to refer to a receiver. The dictator has a **set of actions**, $A_d := \{uneq, egal\}$, with typical element a_d. In this game, the set of actions coincides with the strategy space of the dictator. The receiver has no strategy in this game. We define the

payoff function of the dictator (resp. receiver) as a mapping from the strategy space of the dictator to the set of real numbers. More formally, $\pi_d : A_d \to \mathbb{R}$ (resp. $\pi_r : A_d \to \mathbb{R}$). To represent the payoff functions of both players analytically, let us first define k to be an arbitrarily large number, and the largest of all considered; $\bar{x}$ to be another arbitrarily large number, but smaller than the former; $\tilde{x}$ to be an arbitrarily medium number; and $\underline{x}$ to be an arbitrarily small number, all lying within the set of real numbers. Thus, we effectively impose the restrictions $k > \bar{x} > \tilde{x} \geq \underline{x}$ to the feasible range of values of those payoffs. When the players face **gains**, we can represent their respective payoff functions as

$$\pi_d^+ := \begin{cases} \bar{x} & \text{if } A_d = uneq \\ \tilde{x} & \text{if } A_d = egal \end{cases}$$

$$\pi_r^+ := \begin{cases} \underline{x} & \text{if } A_d = uneq \\ \tilde{x} & \text{if } A_d = egal \end{cases}$$

However, when they face **losses**, we can represent their payoff functions as

$$\pi_d^- := \begin{cases} \bar{x} - k & \text{if } A_d = uneq \\ \tilde{x} - k & \text{if } A_d = egal \end{cases}$$

$$\pi_r^- := \begin{cases} \underline{x} - k & \text{if } A_d = uneq \\ \tilde{x} - k & \text{if } A_d = egal \end{cases}$$

We define all the **potential combinations of payoffs** of both players when facing gains (resp. losses) with the cartesian product $\pi_d^+ \times \pi_r^+$ (resp. $\pi_d^- \times \pi_r^-$). We define $\langle x_d, x_r \rangle$ as a typical element of either set; where the first element of the ordered pair refers to the payoff of the dictator and the second element refers to the payoff of the receiver. The **utility function** of a generic player $i \in \mathcal{I}$ when facing gains (resp. losses) is then defined as a mapping from the relevant cartesian product to the real number space: $U_i^+ : \pi_d^+ \times \pi_r^+ \to \mathbb{R}$ (resp. $U_i^- : \pi_d^- \times \pi_r^- \to \mathbb{R}$). We define T, with typical element t, as the set of potential **theories** of utility that we consider in this paper. For the remainder of the appendix, and for compactness of the proofs, we will refer to the utility function of theory t simply as $U_{i,t}$, which is a step function that takes the form $U_{i,t}^+$ when facing gains and the form $U_{i,t}^-$ when facing losses.

Normal Form Representation. We can, now, define a **binary dictator game** over gains (resp. losses) where subjects are endowed with a generic preference profile t as Γ_t^+ (resp. Γ_t^-). A dictator game over gains (resp. losses) is represented by its set of players, the set of actions of the dictator, and the utility functions of each player. Formally, we write $\Gamma_t^+ := \{\mathcal{I}, A_d, U_{d,t}^+, U_{r,t}^+\}$ (resp. $\Gamma_t^- := \{\mathcal{I}, A_d, U_{d,t}^-, U_{r,t}^-\}$) to denote the normal form

representation of a dictator game over gains (resp. losses) where players are endowed with preference profiles informed by theory t.

Theories. Finally, let us formally present the dictator's utility functions of the theories that we consider in this paper below

$$U_{d,he} := \begin{cases} \pi_d^+ & \text{if } \Gamma_{he}^+ \\ \pi_d^- & \text{if } \Gamma_{he}^- \end{cases} \tag{12}$$

$$U_{d,ia} := \begin{cases} \pi_d^+ - \beta_d \cdot \max(\pi_d^+ - \pi_r^+, 0) & \text{if } \Gamma_{ia}^+ \\ \lambda_d \cdot \pi_d^+ - \beta_d \cdot \max(\pi_d^+ - \pi_r^+, 0) & \text{if } \Gamma_{ia}^- \end{cases} \tag{13}$$

$$U_{d,se} := \begin{cases} (1-\rho_d) \cdot \pi_d^+ + \rho_d \cdot \sum_{i \in \mathcal{I}} \pi_i^+ & \text{if } \Gamma_{se}^+ \\ \lambda_d \cdot (1-\rho_d) \cdot \pi_d^- + \rho_d \cdot \sum_{i \in \mathcal{I}} \pi_i^- & \text{if } \Gamma_{se}^- \end{cases} \tag{14}$$

$$U_{d,mm} := \begin{cases} (1-\gamma_d) \cdot \pi_d^+ + \gamma_d \cdot \min\{\pi_d^+, \pi_r^+\} & \text{if } \Gamma_{mm}^+ \\ \lambda_d \cdot (1-\gamma_d) \cdot \pi_d^- + \gamma_d \cdot \min\{\pi_d^-, \pi_r^-\} & \text{if } \Gamma_{mm}^- \end{cases} \tag{15}$$

The utility function 12 captures an individual that only cares about their material payoff, laying the foundation for our folk understanding of selfishness (viz., homo economicus, hence the notation $t = he$). The utility function represented by equation 13 captures a broader notion of selfishness, where an individual only cares about their own utility, but now this utility also incorporates some personal dislike of self-centered advantageous inequality. This is a representation of inequality aversion preferences (hence the notation $t = ia$), where we omit disadvantageous inequality as in the games we focus on the dictator is always at least as well off as the receiver. In equation 13, the parameter $0 \leq \beta_d < 1$ captures the strength of the dictator's disutility from having a higher payoff than the receiver. The utility function 14 represents the utility function of a dictator that cares about their own material payoff and the aggregate material payoff of the players involved in a game (viz., social efficiency, hence the notation $t = se$). The parameter $0 \leq \rho_d \leq 1$ measures the dictator's degree of concern for the aggregate material payoff of the players in a game. Equation 15 represents the utility function of a dictator who cares about their own material payoff and the minimum payoff achieved in a given game (viz., maximin, hence the notation $t = mm$). Again, the parameter $0 \leq \gamma_d \leq 1$ measures the dictator's degree of concern for the motivation (in this case, maximin). Finally, the parameter λ_d in equations 13, 14, and 15 captures a subject's subjective distortion of one's own material payoff when in losses. Whenever $\lambda_d > 1$, a subject displays loss aversion in the sense that they give more weight to their own material payoff relative to the relevant social motivation they care about. Whenever $\lambda_d = 1$, a subject's strength of like of their material payoff relative to the social motive they care about is the same over gains and losses. Whenever $\lambda_d < 1$, a subject's strength of like of their own material payoff relative to the relevant social motivation decreases when facing losses.

Inequality Aversion

PROPOSITION 1 (Inequality Aversion). Choosing $a_d = egal$ in Γ^+_{ia} reveals $\beta_d > \frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}$ and choosing $a_d = egal$ in Γ^-_{ia} reveals $\beta_d > \lambda_d \cdot \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$.

Proof of proposition 1. *Step* 1.—If a dictator chooses $a_d = egal$ in game Γ^+_{ia} whenever $a_d = uneq$ is present, and the choice is driven by a strict preference, then it follows that $U^+_{d,ia}(a_d = egal) > U^+_{d,ia}(a_d = uneq)$. We substitute the functional form of the utilities to solve the inequality for the only parameter, β_d, as follows

$$
\begin{aligned}
U^+_{d,ia}(egal) > U^+_{d,ia}(uneq) \Leftrightarrow & \\
& \Leftrightarrow \tilde{x} > \bar{x} - \beta_d \cdot (\bar{x} - \underline{x}) && \downarrow \textit{Substitute (13)} \\
& = \beta_d \cdot (\bar{x} - \underline{x}) > \bar{x} - \tilde{x} && \downarrow \textit{Isolate } \beta_d \\
& = \beta_d > \tfrac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}} && \downarrow \div (\bar{x} - \underline{x})
\end{aligned}
$$

Step 2.—If a dictator chooses $a_d = egal$ in game Γ^-_{ia} whenever $a_d = uneq$ is present, and the choice is driven by a strict preference, then it follows that $U^-_{d,ia}(a_d = egal) > U^-_{d,ia}(a_d = uneq)$. We substitute the functional form of the utilities to solve the inequality for the parameter β_d in terms of λ_d as follows

$$
\begin{aligned}
U^-_{d,ia}(egal) > U^-_{d,ia}(uneq) \Leftrightarrow & \\
& \Leftrightarrow \lambda_d \cdot (\bar{x} - k) > \lambda_d \cdot (\bar{x} - k) - \beta_d \cdot \left(\bar{x} - \boxed{k} - \underline{x} + \boxed{k}\right) && \downarrow \textit{Substitute (13)} \\
& = \beta_d \cdot (\bar{x} - \underline{x}) > \boxed{\lambda_d} \cdot (\bar{x} - k) - \boxed{\lambda_d} \cdot (\bar{x} - k) && \downarrow \textit{Isolate } \beta_d \textit{ \& simplify} \\
& = \beta_d \cdot (\bar{x} - \underline{x}) > \lambda_d \cdot \left(\bar{x} - \boxed{k} - \tilde{x} + \boxed{k}\right) && \downarrow \textit{Simplify} \\
& = \beta_d \cdot (\bar{x} - \underline{x}) > \lambda_d \cdot (\bar{x} - \tilde{x}) && \\
& = \beta_d > \lambda_d \cdot \left(\tfrac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right) && \downarrow \div (\bar{x} - \underline{x})
\end{aligned}
$$

■

COROLLARY 1.1 (Inequaliy Aversion and Irrelevance of framing). If $\lambda_d = 1$ (i.e., in the absence of reference dependence), then the dictator's utility maximizing action is the same in both Γ^+_{ia} and Γ^-_{ia}.

Proof of corollary 1.1. We assume that $\lambda_d = 1$ and that the dictator d chooses different actions in Γ^+_{ia} and Γ^-_{ia}. We, then, proceed to prove the corollary by contradiction.
Step 1.—In proposition 1 we have demonstrated that choosing $a_d = egal$ in Γ^-_{ia} reveals $\beta_d > \lambda_d \cdot \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$. If $\lambda_d = 1$, then the threshold becomes $\beta_d > \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$; which is the same as the one revealed when choosing $a_d = egal$ in Γ^+_{ia}.
Step 2.—Now assume that a dictator chooses $a_d = egal$ in Γ^+_{ia} and $a_d = uneq$ in Γ^-_{ia}. In such case, the former reveals $\beta_d > \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$ and the latter reveals $\beta_d < \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$. It is straightforward to see that there is no feasible value of β_d for both restrictions. If we otherwise assume that a dictator chooses $a_d = uneq$ in Γ^+_{ia} and $a_d = egal$ in Γ^-_{ia}, then

the former reveals $\beta_d < \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$ and the latter reveals $\beta_d > \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$. As with the latter assumption, there is no feasible value that fulfills both restrictions. Therefore, and if $\lambda_d = 1$, it must be that either $a_d = egal$ is chosen in both games, which can be sustained by a dictator maximizing $U_{d,ia}$ preferences with values of $\beta_d \in \left[\left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right), 1\right)$; or that $a_d = uneq$ is chosen in both games, which can be sustained by a dictator maximizing $U_{d,ia}$ preferences with values of $\beta_d \in \left[0, \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)\right]$. ■

Social Efficiency

PROPOSITION 2 (Social Efficiency). Choosing $a_d = egal$ in Γ^+_{se} reveals $\rho_d > \frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}$ and choosing $a_d = egal$ in Γ^-_{se} reveals $\rho_d > \lambda_d \cdot \left(\frac{\bar{x}-\tilde{x}}{(2-\lambda_d)\cdot\tilde{x}+(\lambda_d-1)\cdot\bar{x}-\underline{x}}\right)$.

Proof of proposition 2. *Step* 1.—If a dictator chooses $a_d = egal$ in game Γ^+_{se} whenever $a_d = uneq$ is present, and the choice is driven by a strict preference, then it follows that $U^+_{d,se}(a_d = egal) > U^+_{d,se}(a_d = uneq)$. We substitute the functional form of the utilities to solve the inequality for the only parameter, ρ_d, as follows

$$
\begin{aligned}
& U^+_{d,se}(egal) > U^+_{d,se}(uneq) \Leftrightarrow && \\
& \Leftrightarrow (1-\rho_d)\cdot\tilde{x} + \rho_d \cdot 2 \cdot \tilde{x} > (1-\rho_d)\cdot\bar{x} + \rho_d \cdot (\bar{x}+\underline{x}) && \downarrow \textit{Substitute (14)} \\
& = \tilde{x} - \rho_d\cdot\tilde{x} + \rho_d\cdot 2\cdot\tilde{x} > \bar{x} - \rho_d\cdot\bar{x} + \rho_d\cdot\bar{x} + \rho_d\cdot\underline{x} && \downarrow \textit{Expand} \\
& = \rho_d\cdot(\tilde{x}-\underline{x}) > \bar{x}-\tilde{x} && \downarrow \textit{Isolate } \rho_d \\
& = \rho_d > \left(\tfrac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right) && \downarrow \div(\tilde{x}-\underline{x})
\end{aligned}
$$

Step 2.—If a dictator chooses $a_d = egal$ in game Γ^-_{mm} whenever $a_d = uneq$ is present, and the choice is driven by a strict preference, then it follows that $U^-_{d,mm}(a_d = egal) > U^-_{d,mm}(a_d = uneq)$. We substitute the functional form of the utilities to solve the inequality for the parameter γ_d in terms of λ_d as follows

$$
\begin{aligned}
& U^-_{d,se}(egal) > U^-_{d,se}(uneq) \Leftrightarrow && \\
& \Leftrightarrow \lambda_d\cdot(1-\rho_d)\cdot(\bar{x}-k) + \rho_d\cdot(2\cdot\bar{x}-2\cdot k) > \lambda_d\cdot(1-\rho_d)\cdot(\bar{x}-k) + \rho_d\cdot(\bar{x}+\underline{x}-2\cdot k) && \downarrow \textit{Substitute (14)} \\
& = \rho_d\cdot(\bar{x}\cdot(2-\lambda_d) + \bar{x}\cdot(\lambda_d-1) - \underline{x}) > \lambda_d\cdot(\bar{x}-\tilde{x}) && \downarrow \textit{Expand \& Simplify} \\
& = \rho_d > \lambda_d\cdot\left(\tfrac{\bar{x}-\tilde{x}}{\bar{x}\cdot(2-\lambda_d)+\bar{x}\cdot(\lambda_d-1)-\underline{x}}\right) && \downarrow \div(\bar{x}\cdot(2-\lambda_d)\cdots)
\end{aligned}
$$

■

COROLLARY 2.1 (Social Efficiency and Irrelevance of framing). If $\lambda_d = 1$ (i.e., in the absence of reference dependence), then the dictator's utility maximizing action is the same in both Γ^+_{se} and Γ^-_{se}.

Proof of corollary 2.1. We assume that $\lambda_d = 1$ and that the dictator d chooses different actions in Γ^+_{se} and Γ^-_{se}. We, then, proceed to prove the corollary by contradiction.

Step 1.—In proposition 2 we have demonstrated that choosing $a_d = egal$ in Γ^-_{se} reveals $\rho_d > \lambda_d \cdot \left(\frac{\bar{x}-\tilde{x}}{\bar{x}\cdot(2-\lambda_d)+\bar{x}\cdot(\lambda_d-1)-\underline{x}}\right)$. If $\lambda_d = 1$, then the threshold becomes $\rho_d > \left(\frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right)$; which is the same as the one revealed when choosing $a_d = egal$ in Γ^+_{se}.

Step 2.—Now assume that a dictator chooses $a_d = egal$ in Γ^+_{se} and $a_d = uneq$ in Γ^-_{se}. In such case, the former reveals $\rho_d > \left(\frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right)$ and

the latter reveals $\rho_d < \left(\frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right)$. It is straightforward to see that there is no feasible value of ρ_d for both restrictions. If we otherwise assume that a dictator chooses $a_d = uneq$ in Γ^{+}_{se} and $a_d = egal$ in Γ^{-}_{se}, then the former reveals $\rho_d < \left(\frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right)$ and the latter reveals $\rho_d > \left(\frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right)$. As with the latter assumption, there is no feasible value that fulfills both restrictions. Therefore, and if $\lambda_d = 1$, it must be that either $a_d = egal$ is chosen in both games, which can be sustained by a dictator maximizing $U_{d,se}$ preferences with values of $\rho_d \in \left[\left(\frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right), 1\right)$; or that $a_d = uneq$ is chosen in both games, which can be sustained by a dictator maximizing $U_{d,se}$ preferences with values of $\rho_d \in \left[0, \left(\frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}}\right)\right]$. ■

Maximin

PROPOSITION 3 (Maximin). Choosing $a_d = egal$ in Γ^{+}_{mm} reveals $\gamma_d > \frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}$ and choosing $a_d = egal$ in Γ^{-}_{mm} reveals $\gamma_d > \lambda_d \cdot \frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}+\lambda_d\cdot(\bar{x}-\tilde{x})}$.

Proof of proposition 3. *Step* 1.—If a dictator chooses $a_d = egal$ in game Γ^{+}_{mm} whenever $a_d = uneq$ is present, and the choice is driven by a strict preference, then it follows that $U^{+}_{d,mm}(a_d = egal) > U^{+}_{d,mm}(a_d = uneq)$. We substitute the functional form of the utilities to solve the inequality for the only parameter, γ_d, as follows

$$
\begin{aligned}
&U^{+}_{d,mm}(egal) > U^{+}_{d,mm}(uneq) \Leftrightarrow && \\
&\Leftrightarrow (1-\gamma_d)\cdot\tilde{x} + \gamma_d\cdot\tilde{x} > (1-\gamma_d)\cdot\bar{x} + \gamma_d\cdot(\underline{x}) && \downarrow \textit{Substitute (15)} \\
&= \tilde{x} > \bar{x} - \gamma_d\cdot\bar{x} + \gamma_d\cdot\underline{x} && \downarrow \textit{Simplify} \\
&= \gamma_d\cdot(\bar{x}-\underline{x}) > \bar{x}-\tilde{x} && \downarrow \textit{Isolate } \gamma_d \\
&= \gamma_d > \left(\tfrac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right) && \downarrow \div(\bar{x}-\underline{x})
\end{aligned}
$$

Step 2.—If a dictator chooses $a_d = egal$ in game Γ^{-}_{mm} whenever $a_d = uneq$ is present, and the choice is driven by a strict preference, then it follows that $U^{-}_{d,mm}(a_d = egal) > U^{-}_{d,mm}(a_d = uneq)$. We substitute the functional form of the utilities to solve the inequality for the parameter γ_d in terms of λ_d as follows

$$
\begin{aligned}
&U^{-}_{d,mm}(egal) > U^{-}_{d,mm}(uneq) \Leftrightarrow && \\
&\Leftrightarrow \lambda_d\cdot(1-\gamma_d)\cdot(\tilde{x}-k) + \gamma_d\cdot(\tilde{x}-k) > \lambda_d\cdot(1-\gamma_d)\cdot(\bar{x}-k) + \gamma_d\cdot(\underline{x}-k) && \downarrow \textit{Substitute (15)} \\
&= \gamma_d\cdot(\tilde{x}-\underline{x}+\lambda_d\cdot(\bar{x}-\tilde{x})) > \lambda_d\cdot(\bar{x}-\tilde{x}) && \downarrow \textit{Expand \& Simplify} \\
&= \gamma_d > \lambda_d\cdot\left(\tfrac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}+\lambda_d\cdot(\bar{x}-\tilde{x})}\right) && \downarrow \div(\tilde{x}-\underline{x}+\lambda_d\cdot(\bar{x}-\tilde{x}))
\end{aligned}
$$

■

COROLLARY 3.1 (Maximin and Irrelevance of framing). If $\lambda_d = 1$ (i.e., in the absence of reference dependence), then the dictator's utility-maximizing action is the same both Γ^{+}_{mm} and Γ^{-}_{mm}.

Proof of corollary 3.1. We assume that $\lambda_d = 1$ and that the dictator d chooses different actions in Γ^{+}_{mm} and Γ^{-}_{mm}. We, then, proceed to prove the corollary by contradiction.
Step 1.—In proposition 3 we have demonstrated that choosing $a_d = egal$ in Γ^{-}_{mm} reveals $\gamma_d > \lambda_d \cdot \frac{\bar{x}-\tilde{x}}{\tilde{x}-\underline{x}+\lambda_d\cdot(\bar{x}-\tilde{x})}$. If $\lambda_d = 1$, then the threshold becomes $\gamma_d > \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$; which is the same as the one revealed when choosing

$a_d = egal$ in Γ^+_{mm}.

Step 2.—Now assume that a dictator chooses $a_d = egal$ in Γ^+_{mm} and $a_d = uneq$ in Γ^-_{mm}. In such case, the former reveals $\gamma_d > \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$ and the latter reveals $\gamma_d < \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$. It is straightforward to see that there is no feasible value of γ_d for both restrictions. If we otherwise assume that a dictator chooses $a_d = uneq$ in Γ^+_{mm} and $a_d = egal$ in Γ^-_{mm}, then the former reveals $\gamma_d < \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$ and the latter reveals $\gamma_d > \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)$. As with the latter assumption, there is no feasible value that fulfills both restrictions. Therefore, and if $\lambda_d = 1$, it must be that either $a_d = egal$ is chosen in both games, which can be sustained by a dictator maximizing $U_{d,mm}$ preferences with values of $\gamma_d \in \left[\left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right), 1\right)$; or that $a_d = uneq$ is chosen in both games, which can be sustained by a dictator maximizing $U_{d,mm}$ preferences with values of $\gamma_d \in \left[0, \left(\frac{\bar{x}-\tilde{x}}{\bar{x}-\underline{x}}\right)\right]$. ■

References

Alger, Ingela, and Jörgen W. Weibull. 2013. "Homo Moralis—Preference Evolution Under Incomplete Information and Assortative Matching." *Econometrica* 81 (6): 2269–2302. https://doi.org/10.3982/ECTA10637.

Andreoni, James, and B. Douglas Bernheim. 2009. "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77 (5): 1607–1636. https://doi.org/10.3982/ECTA7384.

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (2): 737–753. http://www.jstor.org/stable/2692289.

Bardsley, Nicholas. 2000. "Control without deception: Individual behaviour in free-riding experiments revisited." *Experimental Economics* 3:215–240. https://doi.org/10.1023/A:1011420500828.

———. 2008. "Dictator game giving: altruism or artefact?" *Experimental economics* 11:122–133. https://doi.org/10.1007/s10683-007-9172-2.

Becker, Gary S. 1974. "A Theory of Social Interactions." *Journal of Political Economy* 82 (6): 1063–1093. https://doi.org/10.1086/260265.

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak. 1964. "Measuring utility by a single-response sequential method." *Behavioral Science* 9 (3): 226–232. https://doi.org/10.1002/bs.3830090304.

Bellemare, Charles, Sabine Kröger, and Arthur Van Soest. 2008. "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." *Econometrica* 76 (4): 815–839. https://doi.org/10.1111/j.1468-0262.2008.00860.x.

Benistant, Julien, and Rémi Suchon. 2021. "It does (not) get better: Reference income violation and altruism." *Journal of Economic Psychology* 85:1–21. https://doi.org/10.1016/j.joep.2021.102380.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2012. "What Do You Think Would Make You Happier? What Do You Think You Would Choose?" *American Economic Review* 102 (5): 2083–2110. https://doi.org/10.1257/aer.102.5.2083.

Blanchflower, David G., and Andrew J. Oswald. 2004. "Well-being over time in Britain and the USA." *Journal of Public Economics* 88 (7): 1359–1386. https://doi.org/10.1016/S0047-2727(02)00168-8.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann. 2011. "A within-subject analysis of other-regarding preferences." *Games and Economic Behavior* 72 (2): 321–338. https://doi.org/10.1016/j.geb.2010.09.008.

Bolton, Gary E., Elena Katok, and Rami Zwick. 1998. "Dictator game giving: Rules of fairness versus acts of kindness." *International journal of game theory* 27:269–299. https://doi.org/10.1007/s001820050072.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (1): 166–193. https://doi.org/10.1257/aer.90.1.166.

Boun My, Kene, Nicolas Lampach, Mathieu Lefebvre, and Jacopo Magnani. 2018. "Effects of gain-loss frames on advantageous inequality aversion." *Journal of the Economic Science Association* 4:99–109. https://doi.org/10.1007/s40881-018-0057-2.

Bowles, Samuel, and Sandra Polania-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature* 50 (2): 368–425. https://doi.org/10.1257/jel.50.2.368.

Breitmoser, Yves, and Jonathan H.W. Tan. 2013. "Reference dependent altruism in demand bargaining." *Journal of Economic Behavior & Organization* 92:127–140. https://doi.org/10.1016/j.jebo.2013.06.001.

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer. 2024. "Meta-analysis of Empirical Estimates of Loss Aversion." *Journal of Economic Literature* 62 (2): 485–516. https://doi.org/10.1257/jel.20221698.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. 2018. "The many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences." *Journal of the European Economic Association* 17 (4): 1025–1069. https://doi.org/10.1093/jeea/jvy018.

Camerer, Colin F. 2003. "Dictator, Ultimatum, and Trust Games." In *Behavioral game theory: Experiments in strategic interaction,* edited by Colin F. Camerer and Ernst Fehr, 43–117. The Roundtable Series in Behavioral Economics. Princeton university press. https://press.princeton.edu/books/hardcover/9780691090399/behavioral-game-theory.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3): 818–827. https://doi.org/10.1257/aer.97.3.818.

Cappelen, Alexander W., Ulrik H. Nielsen, Erik Ø. Sørensen, Bertil Tungodden, and Jean-Robert Tyran. 2013. "Give and take in dictator games." *Economics Letters* 118 (2): 280–283. https://doi.org/10.1016/j.econlet.2012.10.030.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." *American Economic Review* 102 (6): 2981–3003. https://doi.org/10.1257/aer.102.6.2981.

Cardenas, Juan Camilo, and Jeffrey Carpenter. 2008. "Behavioural Development Economics: Lessons from Field Labs in the Developing World." *The Journal of Development Studies* 44 (3): 311–338. https://doi.org/10.1080/00220380701848327.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests*." *The Quarterly Journal of Economics* 117 (3): 817–869. https://doi.org/10.1162/003355302760193904.

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren. 2002. "Hardnose the Dictator." *American Economic Review* 92 (4): 1218–1221. https://doi.org/10.1257/000282802 60344740.

Clark, Andrew E., Paul Frijters, and Michael A. Shields. 2008. "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles." *Journal of Economic Literature* 46 (1): 95–144. https://doi.org/10.1257/jel.46.1.95.

Cochard, François, and Alexandre Flage. 2024. "Sharing losses in dictator and ultimatum games: A meta-analysis." *Journal of Economic Psychology* 102:1–20. https://doi.org/10.1016/j.joep.2024.102713.

Cooper, David J., and John H. Kagel. 2016. "Other-regarding preferences." In *The handbook of experimental economics,* edited by John H. Kagel and Alvin E. Roth, 2:217–289. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691139999/the-handbook-of-experimental-economics-volume-2.

Cubitt, Robin P., Michalis Drouvelis, Simon Gächter, and Ruslan Kabalin. 2011. "Moral judgments in social dilemmas: How bad is free riding?" New Directions in the Economics of Welfare: Special Issue Celebrating Nobel Laureate Amartya Sen's 75th Birthday, *Journal of Public Economics* 95 (3): 253–264. https://doi.org/10.1016/j.jpubeco.2010.10.011.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes. 2006. "What you don't know won't hurt me: Costly (but quiet) exit in dictator games." *Organizational Behavior and Human Decision Processes* 100 (2): 193–201. https://doi.org/10.1016/j.obhdp.2005.10.001.

Deaton, Angus, and Arthur A. Stone. 2013. "Two Happiness Puzzles." *American Economic Review* 103 (3): 591–597. https://doi.org/10.1257/aer.103.3.591.

Di Tella, Rafael, Robert J. MacCulloch, and Andrew J. Oswald. 2001. "Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness." *American Economic Review* 91 (1): 335–341. https://doi.org/10.1257/aer.91.1.335.

Diaz, Lina, Daniel Houser, John Ifcher, and Homa Zarghamee. 2023. "Estimating social preferences using stated satisfaction: Novel support for inequity aversion." *European Economic Review* 155:1–39. https://doi.org/10.1016/j.euroecorev.2023.104436.

Eckel, Catherine C., and Philip J. Grossman. 2001. "Are Women Less Selfish Than Men?: Evidence From Dictator Experiments." *The Economic Journal* 108 (448): 726–735. https://doi.org/10.1111/1468-0297.00311.

Engel, Christoph. 2011. "Dictator games: A meta study." *Experimental economics* 14:583–610. https://doi.org/10.1007/s10683-011-9283-7.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation*." *The Quarterly Journal of Economics* 114 (3): 817–868. https://doi.org/10.1162/003355399556151.

Fiedler, Susann, and Adrian Hillenbrand. 2020. "Gain-loss framing in interdependent choice." *Games and Economic Behavior* 121:232–251. https://doi.org/10.1016/j.geb.2020.02.008.

Fishburn, Peter C. 1970. "Utility theory with inexact preferences and degrees of preference." *Synthese* 21 (2): 204–221. https://doi.org/10.1007/BF00413546.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits. 2007. "Individual Preferences for Giving." *American Economic Review* 97 (5): 1858–1876. https://doi.org/10.1257/aer.97.5.1858.

Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior* 6 (3): 347–369. https://doi.org/10.1006/game.1994.1021.

Frey, Bruno S., and Alois Stutzer. 2002. "What Can Economists Learn from Happiness Research?" *Journal of Economic Literature* 40 (2): 402–435. https://doi.org/10.1257/002205102320161320.

Gächter, Simon, Eric J. Johnson, and Andreas Herrmann. 2022. "Individual-level loss aversion in riskless and risky choices." *Theory and Decision* 92 (3): 599–624. https://doi.org/10.1007/s11238-021-09839-8.

Gavassa-Pérez, Ernesto María. 2022. "From morality to rules to choices: introducing and testing a new theory on how morals influence cooperation." PhD diss., University of Nottingham. https://eprints.nottingham.ac.uk/68972/.

Goeree, Jacob K., Margaret A. McConnell, Tiffany Mitchell, Tracey Tromp, and Leeat Yariv. 2010. "The 1/d Law of Giving." *American Economic Journal: Microeconomics* 2 (1): 183–203. https://doi.org/10.1257/mic.2.1.183.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An experimental analysis of ultimatum bargaining." *Journal of Economic Behavior & Organization* 3 (4): 367–388. https://doi.org/10.1016/0167-2681(82)90011-7.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4): 309–321. https://doi.org/10.1086/257678.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review* 91 (2): 73–78. https://doi.org/10.1257/aer.91.2.73.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, et al. 2005. ""Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies." *Behavioral and Brain Sciences* 28 (6): 795–815. https://doi.org/10.1017/S0140525X05000142.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior* 7 (3): 346–380. https://doi.org/10.1006/game.1994.1056.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *The American Economic Review* 86 (3): 653–660. http://www.jstor.org/stable/2118218.

Hutcherson, Cendri A., Benjamin Bushong, and Antonio Rangel. 2015. "A neurocomputational model of altruistic choice and its implications." *Neuron* 87 (2): 451–462. https://doi.org/10.1016/j.neuron.2015.06.031.

Iriberri, Nagore, and Pedro Rey-Biel. 2013. "Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do?" *Quantitative Economics* 4 (3): 515–547. https://doi.org/10.3982/QE135.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98 (6): 1325–1348. https://doi.org/10.1086/261737.

Konow, James. 2003. "Which Is the Fairest One of All? A Positive Analysis of Justice Theories." *Journal of Economic Literature* 41 (4): 1188–1239. https://doi.org/10.1257/002205103771800013.

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11 (3): 495–524. https://doi.org/10.1111/jeea.12006.

Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21 (2): 153–174. https://doi.org/10.1257/jep.21.2.153.

List, John A. 2007. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy* 115 (3): 482–493. https://doi.org/10.1086/519249.

Loewenstein, George F., Leigh Thompson, and Max H. Bazerman. 1989. "Social utility and decision making in interpersonal contexts." *Journal of Personality and Social psychology* 57 (3): 426–441. https://doi.org/10.1037/0022-3514.57.3.426.

Maxwell, Scott E., and Harold D. Delaney. 1993. "Bivariate median splits and spurious statistical significance." *Psychological bulletin* 113 (1): 181–190. https://doi.org/10.1037/0033-2909.113.1.181.

McClelland, Gary H., John G. Lynch, Julie R. Irwin, Stephen A. Spiller, and Gavan J. Fitzsimons. 2015. "Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power." *Journal of Consumer Psychology* 25 (4): 679–689. https://doi.org/10.1016/j.jcps.2015.05.006.

Mosimann, James E. 1962. "On the Compound Multinomial Distribution, the Multivariate β- Distribution, and Correlations Among Proportions." *Biometrika* 49 (1/2): 65–82. http://www.jstor.org/stable/2333468.

Nunnari, Salvatore, and Massimiliano Pozzi. 2022. *Meta-Analysis of Inequality Aversion Estimates.* CESifo Working Paper No. 9851. https://doi.org/10.2139/ssrn.4169385.

Oxoby, Robert J., and John Spraggon. 2008. "Mine and yours: Property rights in dictator games." *Journal of Economic Behavior & Organization* 65 (3): 703–713. https://doi.org/10.1016/j.jebo.2005.12.006.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *The American Economic Review* 83 (5): 1281–1302. http://www.jstor.org/stable/2117561.

Ratcliff, Roger, and Gail McKoon. 2008. "The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks." *Neural Computation* 20 (4): 873–922. https://doi.org/10.1162/neco.2008.12-06-420.

Roemer, John E. 2010. "Kantian Equilibrium." *The Scandinavian Journal of Economics* 112 (1): 1–24. https://doi.org/10.1111/j.1467-9442.2009.01592.x.

Roth, Alvin E. 2007. "Repugnance as a Constraint on Markets." *Journal of Economic Perspectives* 21 (3): 37–58. https://doi.org/10.1257/jep.21.3.37.

Schurter, Karl, and Bart J. Wilson. 2009. "Justice and Fairness in the Dictator Game." *Southern Economic Journal* 76 (1): 130–145. https://doi.org/10.4284/sej.2009.76.1.130.

Sen, Amartya K. 1977. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Philosophy & Public Affairs* 6 (4): 317–344. http://www.jstor.org/stable/2264946.

Shaftesbury, Lord. 2000. "An inquiry concerning virtue or merit." In *Shaftesbury: Characteristics of Men, Manners, Opinions, Times,* edited by Lawrence E. Klein, 163–230. Cambridge Texts in the History of Philosophy. Cambridge University Press. https://doi.org/10.1017/CBO9780511803284.

Smith, Vernon L., and Bart J. Wilson. 2015. "Fair and impartial spectators in experimental economic behavior: using sympathy to drive action." In *Sympathy: A History,* edited by Eric Schliesser, 359–385. Oxford Philosophical Concepts. Oxford University Press, USA. https://doi.org/10.1093/acprof:oso/9780199928873.001.0001.

———. 2017. "Sentiments, Conduct, and Trust in the Laboratory." *Social Philosophy and Policy* 34 (1): 25–55. https://doi.org/10.1017/S0265052517000024.

———. 2019. *Humanomics: Moral sentiments and the wealth of nations for the twenty-first century.* Cambridge University Press. https://doi.org/10.1017/9781108185561.

Sobel, Joel. 2005. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature* 43 (2): 392–436. https://doi.org/10.1257/0022051054661530.

Sugden, Robert. 1982. "On the Economics of Philanthropy." *The Economic Journal* 92 (366): 341–350. https://doi.org/10.2307/2232444.

———. 1984. "Reciprocity: The Supply of Public Goods Through Voluntary Contributions." *The Economic Journal* 94 (376): 772–787. https://doi.org/10.2307/2232294.

Weir, B. S., and W. G. Hill. 2002. "Estimating F-Statistics." *Annual Review of Genetics* 36:721–750. https://doi.org/10.1146/annurev.genet.36.050802.093940.

Yu, Peng, and Chad A. Shaw. 2014. "An efficient algorithm for accurate computation of the Dirichlet-multinomial log-likelihood function." *Bioinformatics* 30 (11): 1547–1554. https://doi.org/10.1093/bioinformatics/btu079.