STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Fall 2025

The Structure of Fairness Preferences: Revisiting the Pluralism of Fairness Ideals

Love Leijonklo (24851)

Abstract: The field of normative economics has emerged to challenge assumptions made in traditional microeconomics. However, canonical articles in the field has also themselves produced assumptions that are not well discussed. In this thesis I revisit the experiment of Cappelen (2007), and test their quadratic loss assumption by replacing the quadratic penalty with a more flexible constant-elasticity-of-substitution (CES) penalty while keeping the same latent type approach. I re-estimate the original experiment and find that pluralism is robust to the alternative specification. Allowing CES curvature yields broadly similar conclusions but shifts some mass across ideals: strict egalitarianism remains the biggest group, but the libertarian ideal overtakes liberal egalitarianism as the second biggest ethical ideal in the data.

Keywords: Social preferences, Distributional preferences, Normative Pluralism, Inequality aversion

JEL: C91, D31, D63

Supervisor: Patrick Agte & Paolo G. Piacquadio
Date submitted: December 7th, 2025
Date examined: December 15th, 2025
Discussants: Niklas Schüler
Examiner: Elena Paltseva

Contents

1 Introduction **2**

2 Literature Review **5**

2.1 Theories of fairness and distributive justice 5

2.2 Evolution of methodology to study fairness 7

2.2.1 The Ultimatum Game 7

2.2.2 Dictator game 8

2.2.3 Games with production 9

2.2.4 One-shot vs Multi-shot 9

2.3 Models of fairness preferences 10

2.3.1 The inequity-aversion model 11

2.3.2 Relative-share model of fairness 11

2.3.3 Social-welfare model of fairness 12

2.4 Role of luck and effort in Distributional Preferences 13

3 Methodology **15**

3.1 Experimental design & data 15

3.2 The quadratic baseline model 16

3.3 The CES alternative model 19

3.4 Structural model and estimation 21

3.4.1 Estimation by Maximum Likelihood 21

3.4.2 Estimating via Efficiency Losses 22

4 Results and Discussion **24**

4.1 Interpretation 26

4.2 Limitation 27

5 Conclusion **28**

6 Appendix **31**

Directory of Aids **32**

1 Introduction

We encounter questions of fairness in everyday life situations, and they enter many discussions on how to shape government and policy best. At the highest level, people hold differing views about what constitutes a fair distribution of wealth in society, often influenced by their philosophical and political leanings. Some regard equality of outcomes as the most just solution, regardless of how those outcomes were generated. Others believe fairness should mean that individuals get to keep exactly what they produce, without the interference of redistribution or high taxes. Between these positions lies a spectrum of views: some emphasise that rewards should be proportional to effort, while others stress that entitlements depend on whether outcomes result from personal choice or chance.

At the heart of these questions is whether individuals should be rewarded for how hard they try, or for what they actually achieve, even when part of the variation of outcome perhaps simply depends on factors of chance. It is therefore of interest to see how fairness judgments depend on how people weigh the relative roles of effort and luck in shaping outcomes. When governments design taxes or welfare systems, they implicitly answer this fairness question: should redistribution be correct for bad luck, low effort, or both? Different societies resolve this tension in other ways, not only in how much they redistribute, but also in how they justify doing so.

While no single system will likely satisfy everyone, it is reasonable to argue that redistributive policies in a democracy ought to align with the ethical views and distributional preferences of the people in the society. Understanding how people reason about fairness, is therefore an important condition for designing political-economic institutions that are publicly legitimate.

How the field of economics can lend itself to capturing fairness views and an economic theory that mathematically models the trade-off between efficiency and fairness is not obvious. And for a long time, the consensus view was that economics should not try. This was during a period when welfare economics was in the utilitarian tradition, which states that social welfare should be assessed by the sum of individual utilities (Bentham, 1789). This approach provided analytical clarity, but also narrowed the range of ethical theories that could be considered. For example, Arrow's impossibility theorem (Arrow, 1951) showed that coherent social preferences cannot be derived from individual rankings without restrictive

assumptions. For decades, this discouraged economists from engaging directly with normative questions such as fairness. With the rise of behavioural economics and the growing use of experiments and surveys, the interest in studying ethical diversity and pluralism has returned. A body of experimental evidence has been generated in the last decades that points to people not only caring about outcomes but also about how those outcomes came to be, whether they were earned, deserved, or the product of luck (Konow, 2000; Cappelen et al., 2007). This has fueled new debates on distributive justice, turning the economics of fairness ideals into a topic of repeated discussion.

In this thesis, I examine one way of separating luck and effort in fairness judgments by re-analysing data from an experimental approach developed in the behavioural economics literature. I build on the study "Pluralism of Fairness Ideals" by Cappelen et al. (2007), which uses a modified dictator game to estimate how individuals trade off fairness against self-interest and how they differ in the fairness ideals they apply. This framework provides a way to study how people think about redistribution when both personal choices and randomness affect outcomes.

The motivation for revisiting the original quadratic specification is that it constrains how fairness concerns are allowed to vary, in other words, it makes an assumptions that are not discussed anywhere in the original article. Perhaps the most straightforward example of the assumptions practical relevance is that of taxation. If the model understates how strongly people dislike large unfairness gaps, it could suggest that people are comfortable with flatter taxes even when, in reality, they could in fact support stronger redistribution. Also, the opposite is true, if the model overstates fairness sensitivity, it might suggest too steep tax rates or high transfers that reduce efficiency. Allowing curvature in the CES model shows whether people react strongly or mildly to inequality, which is important for deciding how progressive taxes and social insurance should be.

Since being published, alternative models have been put forward. My contribution is to test an alternative functional form of utility that other authors in welfare economics propose, on the same data. I thus contribute to the literature by assessing whether the three fairness ideals put forward in the original article are better modelled using an alternative specification and if the shares of the ideals remain robust in such a specification.

The alternative functional forms of utility tested are a ratio-based CES preference around deviation of the fair share, as opposed to a quadratic penalty based on level distance penalty. I find that the fairness ideals proposed in the original article remain robust under the alternative utility function, and argue why the alternative specification aligns more closely

with intuitive notions of fairness and with the properties such preferences should naturally exhibit.

The rest of the thesis is organised as follows: the second section reviews the body of literature on the topic, the different experiments used to measure, theories of distributive justice, and factors affecting social preferences. The third section describes the experimental design and dataset generated. It defines the baseline model used in the original paper, the models tested in this thesis, and the estimation approach employed. The fourth section presents the results and discussion from this analysis, and the fifth section concludes.

2 Literature Review

Understanding how people think about fairness and redistribution requires both theoretical models and empirical observation. Progress in this field has emerged from a continuous interplay between the two: models define the variables of interest and suggest what should be measured. At the same time, experiments test these models and often generate new insights that refine or challenge existing assumptions. Through this feedback loop, knowledge advances incrementally, bit by bit. This interaction between theory and evidence lies at the heart of normative and behavioural economics (Kagel & Roth, 1995). In the study of distributive justice, it has been particularly fruitful: theoretical models clarify what fairness might mean in abstract terms, while experiments reveal how people actually reason about fairness in concrete situations.

This literature review begins with a brief overview of central theories of distributive justice from sociology and political philosophy that have influenced economic thinking on fairness. Next, I turn to the development of experimental approaches to fairness, describing how experimental designs have evolved within behavioural economics and outlining some of the canonical experiments. Following this, I discuss how these empirical insights have been formalised into mathematical models of social preferences. To separate the theoretical and empirical advances is not easy since they are so intertwined, and the second and third sections of the review naturally overlap, as many models were directly informed by experimental findings, and new experiments were often designed to test or distinguish between competing theoretical approaches. Lastly, I turn my attention to the literature on how luck and effort shape people's distributive preferences.

2.1 Theories of fairness and distributive justice

Questions of what counts as a fair distribution are as old as philosophy itself. From the Greek philosophers like Aristotle's reflection on proportionality to modern theories of justice, fairness has been a central concern in ethics and political thought. The topic from a philosophical perspective is boundless, but in the economics literature of fairness the focus is set on a smaller number of fairness ideals that can be translated into clear behavioural predictions. Three ideals are used in the Cappelen study and will be considered: strict egal-

itarianism, libertarianism, and liberal egalitarianism (Cappelen, 2007). They differ in how they weigh equality, individual responsibility, and the role of luck. They are well-suited ideals in a study, since each offers a precise and mutually exclusive answer to what constitutes a fair distribution when outcomes depend on both choice and chance. In this subsection, I briefly cover the meaning of each one of them.

Strict egalitarianism holds that fairness requires equal outcomes for all individuals, regardless of effort, ability, or circumstance. Inequalities are seen as inherently unjust, even if they arise from choices. In experimental design, this principle implies that participants should prefer equal splits of income, no matter how the income was generated. It represents the equality of outcome view of justice, prioritising equality over merit or efficiency.

Libertarianism, by contrast, emphasises entitlements and rights. Under this view, each person keeps the full income that he or she produces, even if the size of that income can be attributed to effort or luck. Inequalities are therefore legitimate as long as they result from voluntary actions and legitimate processes. Unlike the strict egalitarian perspectives, libertarianism focuses on procedural justice, how outcomes arise, rather than the final distribution and relative standing of individuals.

Liberal egalitarianism lies somewhere between the two positions mentioned previously. It holds that a person should bear the consequences of choices for which he or she is responsible, but should not bear the consequences of facts that lie beyond personal control. A difference in income is fair when it comes back to extra effort, it is unfair when it traces back to brute luck. This position links responsibility with fairness but draws a boundary between justifiable and unjustifiable inequalities.

Beyond these three empirically studied fairness ideals, normative economics has also been shaped by broader philosophical principles such as Rawls's maximin rule and Benthamite utilitarianism, which define alternative views of what fairness or social welfare might mean (Niehans, 1990). These principles judge distributions by their moral desirability: maximin focuses on the worst-off individual, while utilitarianism points to maximizing the total welfare. Another benchmark is Pareto efficiency, which does not state what is fair but what is waste-free: an allocation is efficient if no one can be made better off without making someone else worse off. The concept of Pareto efficiency reappears later in the thesis in the evaluation of the models.

For the purpose of empirical identification, however, the focus in experimental economics has been on strict egalitarianism, libertarianism, and liberal egalitarianism. Importantly,

the three ideals generate mutually exclusive fairness prescriptions whenever outcomes reflect both investment and luck. Together, they provide a clear set of normative benchmarks for analyzing distributive preferences. The next step is to consider how these theoretical ideals have been explored in practice, which brings us to the evolution of experimental approaches to fairness.

2.2 Evolution of methodology to study fairness

The study of fairness in economics has long been shaped by methodological debates about how preferences can and should be measured. In the neoclassical school, preferences were not thought to be observable directly through questions, but had to be inferred from observed choices. Paul Samuelson argued that only actual decisions under constraints could reliably uncover preferences (Samuelson, 1947). Preferences over fairness, however, are not the kind of preferences that can easily be observed from market data, since they do not typically manifest themselves in observable consumption or labour choices. This makes distributive preferences difficult to study as revealed preferences, and helps explain why fairness was long neglected in mainstream economics.

Survey-based approaches, such as vignette surveys, offer another way forward but are problematic for a number of reasons. Framing effects often produced large variation in responses, and the absence of real consequences made them vulnerable to framing bias. On the other hand, the value of surveys compared to lab experiments is in the possibility of being scaled to large samples, since they are less costly, and they therefore serve a value as an additional tool.

The change came in the 1980s, when researchers began to pay subjects real money and conduct laboratory experiments where participants face allocation problems to draw out distributive choices. Because of the controlled environment and the fact that every choice affected the participant's own earnings, people took the tasks seriously, and the data became reliable. I describe the main designs in roughly the order they appeared and highlight how each one added a new way to test ideas about fairness.

2.2.1 The Ultimatum Game

The Ultimatum Game, (Güth, Schmittberger & Schwarze 1982), is widely regarded as the first canonical experiment to challenge the standard economic model of self-interested agents. Before its introduction, most economic theory assumed that individuals maximise only their

own payoff, and systematic experimental tests of fairness were virtually nonexistent in economics at the time. The logic of the game is straightforward: two players interact, one as the proposer who suggests how to split a sum of money (for example, SEK 100), and the other as the responder who can either accept or reject the offer. If the responder accepts, both players are paid according to the proposal; if the responder rejects, both receive nothing. Standard economic theory predicts that the proposer should keep almost the entire amount, since any positive offer is strictly better than nothing for the responder and should therefore be accepted. In real experiments, responders often turn down small offers, even though they receive nothing in return. Observers take this as proof that people will give up money to penalize treatment they see as unfair. The Ultimatum Game showed that concerns about fairness shape choices and that finding produced models built on dislike for unequal outcomes, like the inequity aversion model. Yet the same design cannot tell whether the proposer's offer come from a true wish to be fair or from a wish to avoid a refusal. Because of that unclearity, researchers created a new design, the Dictator Game.

2.2.2 Dictator game

The Dictator Game, introduced in the early 1990s, changed the ultimatum game by taking away the responder's right to refuse the offer. The goal was to separate fairness motives from strategic concerns and to see whether positive offers would still occur when rejection was not possible anymore. In this design, one player, the "dictator," decides unilaterally how to allocate a sum of money between herself and another participant. Standard theory predicts that dictators should keep the entire amount, as there are no incentives to share. Yet experimental evidence consistently shows that many dictators transfer positive amounts, often around 20 per cent of the pie (Forsythe, Horowitz, Savin, & Sefton, 1994).

The result shows that people do not base their decisions about sharing only on direct personal gain. Proposed explanations include altruism, fairness concerns, and social image motives. Hoffman et al. (1996), for example, demonstrate that anonymity strongly affects giving: transfers decrease when decisions are more private, indicating that social image concerns play an important role. Other motives could be pure altruism, that the giver values the recipient's welfare, or warm glow giving, the giver values the act of giving itself. It also paved the way for new experimental designs incorporating production, merit, and entitlement, which allow researchers to examine how fairness judgments depend on the origins of income.

2.2.3 Games with production

A common criticism of the standard dictator and ultimatum games is that participants receive their endowments as a "gift from above," without effort or cost. This fact raises the question about external validity, how realistic is it to study distributive behavior in settings where resources are handed out freely? In real life, people often view income as something they have earned, and attitudes toward redistribution may look very different when allocations concern hard-earned resources rather than windfall gains. The artificial nature of windfall endowments also makes behaviour sensitive to framing, which complicates interpretation.

To address these concerns, researchers began introducing production tasks before the distribution phase. In these designs, participants first perform a real-effort task (such as transcription, adding sequence of numbers, etc) or make investment decisions, sometimes under different rates of return, which then determine the size of the available pie in the distributional phase. This opens the possibility to study how distributional choices are affected by effort/merit as well as making the setting feel more realistic.

The introduction of production has important consequences. Early studies showed that participants were less willing to share when endowments had been earned through effort rather than given freely (Cherry, 2002). Other work showed that fairness judgments became sensitive to the source of inequality: people distinguished between outcomes caused by individual decisions and outcomes due to external circumstances (Konow, 2000). These results suggested that fairness is not only about the final distribution but also in the way it was generated. The introduction of production-based games, therefore, marked a major step in the advancement of fairness experiments. Section 2.4 will examine in greater detail the distinction between effort and luck that those designs introduced.

2.2.4 One-shot vs Multi-shot

Another important dimension of experimental design is whether interactions are structured as one-shot or repeated. In a one-shot game, participants make a single allocation decision without the opportunity for future interaction. This eliminates the possibility of strategising based on future expectations and provides a more precise measure of immediate distributive preferences. By contrast, repeated or multi-shot games create scope for strategic considerations. Participants may act more fairly or cooperatively, not purely out of intrinsic fairness motives, but because they anticipate that today's behaviour could affect how others respond in later rounds. Repeated contact allows reputations to form, opens up the possibility for

reciprocal acts and lets people punish stingy partners. Those mechanisms keep cooperation plus sharing alive even among people who possibly only care for their own payoff (Fehr Charness, 2024).

The difference between one-shot and repeated contact is central to interpreting what fairness means. A single round anonymous setup removes away strategic reasons to appear fair. Under those conditions, generous splits point to an internal preference for fairness. Most fairness studies, therefore, use one-shot, anonymous, paid experiments. Those setups remove confounding details such as reputation or expected future profit and give a conservative test of whether fairness motives exist.

Yet this clarity comes with a cost. One-shot designs heighten internal validity, but they remove social interaction which is a large driver of behaviour in the real world, outside the lab. People usually deal with the same partners repeatedly but also care about their standing in the group. The design choice, therefore, creates a tension. A one-shot game exposes raw preferences, whereas a repeated game shows how people behave when interaction history and reputation matter.

Taken together, the evolution of experimental designs shows how the field progressively refined the tools for studying distributive preferences. The Ultimatum Game showed that people care about fairness. The Dictator Game made that point clearer by removing strategy, and later versions added effort so we could see how merit mattered. Finally, the distinction between one-shot and repeated play highlighted the trade-off between identifying intrinsic fairness motives and capturing richer social dynamics. These methodological advances established a robust experimental foundation and raised new puzzles about how to interpret observed behaviour. To address these puzzles, economists developed formal models of fairness preferences, which provide a structured way to link experimental evidence to underlying utility functions.

2.3 Models of fairness preferences

The experiments above motivates a departure from the purely self-interested choice model of traditional microeconomics. Proposing new models that account for fairness is the next challenge, that we now turn to. The starting point is that the utility functions from microeconomics are extended to include these dimensions. Under familiar regularity conditions (e.g., completeness and transitivity, with continuity and monotonicity as appropriate), preferences admit a utility representation. The key step is to let utility depend not only on one's

own payoff but also on others' outcomes. This keeps the problem easy to handle and lets different explanations be compared in a consistent way. I review three models that describe fairness through utility.

2.3.1 The inequity-aversion model

Fehr and Schmidt (1999) proposed one of the most influential models of social preferences, motivated by results from the ultimatum game. The central assumption is that individuals dislike unequal outcomes. Utility depends not only on one's own payoff but also on differences in payoffs relative to others. The utility of individual i is modelled as

$$U_i(x_i, x_j) = x_i - \alpha_i \max(x_j - x_i, 0) - \beta_i \max(x_i - x_j, 0),$$

where x_i and x_j denote the payoffs of player i and j, respectively. The parameters α_i and β_i capture aversion to disadvantageous and advantageous inequality, with the typical assumption that $\alpha_i > \beta_i$. This asymmetry reflects the empirical finding that individuals are more sensitive to being worse off than to being better off than others. The Fehr-Schmidt model provides a parsimonious explanation for why low offers are rejected in ultimatum games, why dictators give positive amounts, and why cooperation in public goods games is often observed despite incentives to free-ride. It also links naturally to the strict egalitarian ideal discussed earlier, as it formalises the dislike for unequal outcomes.

If individual i's payoff is larger than j's, the first term disappears and utility decreases by β_i for every additional unit above j. Conversely, if i earns less than j, the second term disappears and utility decreases by α_i for every unit below j. Since typically $\alpha_i > \beta_i$, the model assumes that the pain of being worse off is stronger than the discomfort of being better off.

The model is closely related to the strict egalitarian view, as both assume that unequal outcomes are intrinsically undesirable. The inequality-aversion framework can therefore be seen as a parametric version of the egalitarian fairness ideal.

2.3.2 Relative-share model of fairness

The ERC (Equity, Reciprocity and Competition) model is based on a similar intuition, but puts emphasis on the relative shares rather than bilateral payoff differences. An individual's utility depends on both their material payoff and their share of the total payoff:

$$U_i = f(x_i, \sigma_i), \qquad \sigma_i = \frac{x_i}{\sum_j x_j}\,.$$

Utility is maximised when all players receive equal shares ($\sigma_i = \frac{1}{n}$). Unlike the Fehr–Schmidt model, ERC is not restricted to pairwise comparisons but generalises to settings with more than two players. It also provides a game-theoretic interpretation of fairness and how relative standing influences competitive behaviour. The main difference from Fehr-Schmidt is that ERC disregards the identity of particular rivals, each person compares his or her own share to the average for the entire group, not to the amount received by one other individual.

2.3.3 Social-welfare model of fairness

Charness and Rabin took the analysis further in 2002, when they proposed a model that combines self-interest with social welfare concerns. Utility is written as a weighted average of own payoff and a social welfare function:

$$U_i = (1 - \lambda_i)x_i + \lambda_i\Big[\delta_i \min_j(x_j) + (1 - \delta_i)\sum_j x_j\Big]\,.$$

The parameter λ_i captures the overall weight given to others' payoffs. δ_i captures the concern for the least well-off relative to overall efficiency. This framework allows for a wide spectrum of fairness motives: at one extreme, individuals may behave in a Rawlsian maximin way by focusing on the worst-off ($\delta_i \to 1$). On the other hand, they can prioritise efficiency and maximise the total surplus ($\delta_i \to 0$). The model therefore accommodates both egalitarian and utilitarian motives, and it can explain why some individuals prioritise helping the least advantaged while others sacrifice income to increase total group payoffs.

By varying the weights, the model spans a wide spectrum of fairness motives: from pure self-interest ($\lambda = 0$) to utilitarian efficiency concerns ($\delta \to 0$) to Rawlsian maximin preferences ($\lambda = 1$, $\delta = 1$). This flexibility lets the model capture cross-subject heterogeneity in the weights placed on efficiency vs. concern for the worst-off.

Later CES-based models (Andreoni Miller, 2002; Fisman, Kariv Markovits, 2007) further generalised this approach by allowing a continuous elasticity of substitution between self-interest and concern for others, so dispersion arises both in weights and in curvature.

2.4 Role of luck and effort in Distributional Preferences

As discussed previously, production-based games increased realism by making outcomes endogenous. Later work kept that idea but split effort from luck to see how people decide what is fair when the final score depends partly on what a person does and partly on outside events. This line of research has moved the field from simple identification of inequality aversion toward a more nuanced understanding of distributive preferences and how different scenarios can affect the preferences. Key contributions include Konow's formulation of the accountability principle and Cappelen et al.'s notion of pluralism of fairness ideals, which show that people differ not only in the weight they place on fairness but also in the principles they apply when distinguishing between deserved and undeserved inequalities.

The accountability principle, introduced by James Konow (2000), serves to distinguish fair from unfair inequalities. According to it, people should be held responsible for outcomes that follow from their choices, such as effort, but not for outcomes driven by luck or circumstances beyond their control. In experiments where subjects performed effort tasks and then allocated earnings, Konow found evidence for this principle. Namely, participants tended to divide rewards proportionally when differences reflected effort, but shifted toward equal splits when differences were due to exogenous factors. The study also indicates that fairness judgments are not purely objective but can be influenced by self-serving biases. When facing trade-offs between self-interest and fairness, individuals often reinterpret what counts as fair in ways favourable to themselves. Konow referred to this as the role of cognitive dissonance in distributive choices. By combining fairness ideals with psychological mechanisms, the accountability principle offered one of the first clear frameworks for how people separate effort from luck in allocation decisions and it continues to serve as a foundation for later research on distributive justice.

Oppenheimer et al. (2004) contributed to the literature by an extended version of the inequity aversion model, to situations where income comes from production. Their experiments modified the standard dictator game by having both players generate value through work, thereby creating entitlements. Results showed that many participants based allocations not only on equality but also on "just deserts," rewarding others according to their contribution. This aligns with the idea that fairness is context-dependent: when outcomes are tied to effort, proportionality norms become more salient than simple equality.

Importantly, the study also highlighted the role of credibility and social context. Stan-

dard dictator games often rely on anonymous, context-free transfers, which can generate doubt about whether allocations truly reach recipients. Frohlich et al. showed that adding production and a social context reduced such doubts and demonstrated more systematic fairness-oriented behaviour. Their findings suggest that fairness cannot be understood solely through universal axioms like inequality aversion, but must account for multiple normative principles depending on the environment. By incorporating "just deserts" into formal models, they emphasised that distributive preferences involve plural motives, with effort-based entitlements often weighing heavily in people's judgments.

3 Methodology

The first section provides a description of the experimental design employed, the dataset and variables generated and introduces some of the notations used (3.1). I then outline the baseline quadratic functional form of utility that is used by Cappelen et al. (3.2), followed by developing the alternative form of CES-like utility (3.3). Lastly, the structural estimation model used for identification of the parameters is described (3.4).

3.1 Experimental design & data

The experiment was conducted at the Choice Lab at the Norwegian School of Economics (NHH). A total of 96 students were recruited to participate (mix of undergraduate and graduate), studying business and/or economics at NHH. Subjects were randomly assigned to six laboratory sessions and interacted anonymously via web-based interfaces. Across the experiment, a total of 190 distributional choices were generated.

The experiment is a one-shot dictator framework with a production stage followed by a distribution stage. In the production stage, each participant received an endowment of 300 Norwegian kroner (NOK) and chose how much to invest in two different projects. Investment choices were restricted to three possibilities: $\{0, 100, 200\}$ NOK per project, with the additional constraint that total investments could not exceed the initial 300 NOK. Let $q_i \in \{0, 100, 200\}$ denote the investment made by individual i. investment decisions proxy effort or personal responsibility, since higher investments increase potential income but also reduce immediate consumption.

To introduce heterogeneity in opportunities, participants were randomly assigned to either a low ($a_i = 2$) or a high return ($a_i = 4$) which acted as a multiplier on the investment they decided on. This assignment simulated unequal opportunities: some individuals faced more favourable production conditions than others despite exerting the same effort. Individual production is given by: $y_i = a_i \cdot q_i$ so that income depended both on effort (investment, under individual control) and on luck (the randomly assigned return, outside individual control).

In the distribution stage that followed, each participant was anonymously paired with another participant over the web-interface. Each proposed how to divide the total produced

income between themselves and their partner. Before making their decision, participants were informed about their partner's rate of return, investment decision, and production outcome. One of the two allocation games was then randomly selected to determine final earnings, to guarantee that the distributional choices had real consequences.

The three fairness ideals tested were:

Strict egalitarianism (SE). Fairness requires equal outcomes, independent of effort or luck. If the joint production is $X = y_1 + y_2$, the fair allocation to individual i is

$$m_i^{\mathrm{SE}} = \frac{X}{2}\,.$$

Libertarianism (L). Fairness means that each individual keeps exactly what they produced, regardless of whether differences arose from effort or luck. The fair allocation is

$$m_i^{\mathrm{L}} = y_i = a_i q_i\,.$$

Liberal egalitarianism (LE). Individuals should be held responsible for choices (effort) but not for circumstances beyond their control (luck). The fair allocation is proportional to each individual's share of total investment:

$$m_i^{\mathrm{LE}} = \frac{q_i}{q_1 + q_2}\,X\,.$$

3.2 The quadratic baseline model

Models incorporating fairness concerns treat choices as a trade-off, a higher own monetary payoff raises utility, but deviations from the fairness benchmark impose a penalty or disutility. Let y_i denote individual i own payoff from the allocation and let m_i be the fairness perspective that they follow. The utility of choosing to consume y is written as

$$U_i(y) = y - \beta_i L_i(y; m_i),$$

where $\beta_i \geq 0$ measures the importance of fairness for person i and $L_i(\cdot;\cdot)$ is a loss function that turns deviations from the reference point into disutility. The question at the heart of this thesis is the functional form of $L(\cdot)$, i.e. how exactly deviations from fairness translate into disutility.

The loss is anchored at the ideal $L(m_i; m_i) = 0$, which means that an individual who chooses $y = m_i$ suffers from no penalty. It is non-negative and increases with the absolute deviation $|y - m_i|$, so moving away from the ideal causes the penalty to rise. To capture the idea that larger violations should hurt more than smaller ones, the loss function is usually assumed to be convex, implying increasing marginal disutility the farther y moves from m_i, so that, for instance, keeping slightly more than one's fair share causes mild discomfort, whereas taking most of the pie is strongly punished

Because the monetary scale of problems varies across situations, i.e. the size of the whole pie X differs, normalisation ensures that fairness sensitivity is not mechanically higher simply because the scale is larger. Because the monetary scale of problems varies across situations (the "pie" X differs), a well-behaved loss function should incorporate a simple normalisation of deviations so that the penalty is comparable across states and not mechanically driven by units or scale.

Normalising does not change which options people like more (it preserves the ordinality), it just rescales how big the differences look. This ensures the model does not interpret larger stakes as stronger fairness concerns per definition. In other words, two choices that represent the same proportional departure from the ideal should carry similar penalties even when X differs. In some models, m does not vary, every individual is assumed to follow the same fairness ideal. Such is the structure in the inequity-aversion and ERC models described in Section 2.2, where heterogeneity in normative views is collapsed into a single representative notion of fairness. This is precisely the type of restrictive assumption that economic pluralism seeks to challenge.

In the baseline specification used by Cappelen et al. (2007), the loss takes a quadratic form in level deviations from the fairness ideal, namely:

$$V_i(y; a, q) = \gamma y - \beta_i \frac{\Big(y - m_k(i)(a, q)\Big)^2}{2\,X(a, q)}.$$

This formulation implies a simple geometric structure where utility falls with the square of the distance from one's fairness ideal, scaled by the total available income, implying linearly increasing marginal disutility in the deviation. $m_k(a, q)$ is the ideal implied by type $k \in \{SE, LE, L\}$, $\gamma > 0$ is the marginal utility of own money, and $\beta_i \geq 0$ is individual i's weight on fairness. The penalty is normalised by the term $2X(a, q)$, which scales the deviations to the size of the total pie. Since total output X depends on both the return rate

a and the investment level q, the equation says that different pairs can cause outcomes of different size. This ensures that fairness costs are comparable across problems, the factor 2 merely standardises the quadratic shape so that the penalty is symmetric around the fairness ideal.

Taking the partial derivative of V_i with respect to y and setting it equal to zero gives the first-order condition:

$$\frac{\partial V_i}{\partial y} = \gamma - \frac{\beta_i}{X(a,q)} \left(y - m_k(a,q)\right) = 0,$$

which implies the optimal solution (proposal):

$$y_i^* = m_k(a,q) + \frac{\gamma}{\beta_i} X(a,q).$$

The fairness ideal m_k therefore serves as the behavioural anchor, while the ratio γ/β_i determines how strongly self-interest pulls the choice away from it. A higher β_i (stronger fairness concern) keeps y_i^* closer to the fairness benchmark, whereas a higher γ (valuation of own payoff) shifts the allocation towards oneself. The set-up parallels the Fehr–Schmidt model in some sense, but replaces bilateral payoff comparisons with deviations from a fairness reference point defined by the individual's normative ideal $m_k(a,q)$. Rather than caring directly about what the other player receives, individuals care about how far their own allocation departs from what they perceive as fair, given their fairness type.

The quadratic specification constrains the shape of preferences, disutility is symmetric around m_k and increases with the squared distance from the fairness ideal. If the welfare loss from deviating is flatter near the ideal or steeper in the tails than quadratic, the model can mismeasure β and bias mixture shares λ. In practice, this could mean overstating egalitarianism if subjects react sharply to large deviations, or understating it if they tolerate small departures from equality. The shape of the loss function therefore, directly affects inferred fairness types, since β captures the strength of fairness concerns conditional on that assumed curvature. Finally, the quadratic form links local and distant deviations through the same curvature, leaving no flexibility for losses that behave non-linearly.

3.3 The CES alternative model

Under the quadratic specification, the pain from moving away from the fairness ideal grows at an linearly increasing rate: each extra step of unfairness adds more marginal disutility the further one already is from the ideal. However, the quadratic form imposes a constant curvature, i.e. a constant second derivative, which fixes how quickly marginal disutility increases from small vs large deviations.

To introduce flexibility, I replace the quadratic loss with a ratio-based Constant Elasticity of Substitution (CES) loss function. This form lets the curve bend more or less through a parameter, that fixes how fast disutility grows when the share drifts away from the fair share. It gives a simple test of whether the spread of fairness ideals changes when we change the shape of the loss function.

Formally, I specify individual utility as

$$V_i(y) = \gamma\, y \;-\; \frac{\beta_i}{X}\left[\frac{m_k}{1+\rho}\left(\left(\frac{y}{m_k}\right)^{1+\rho} - 1\right) \;+\; \frac{X-m_k}{1+\rho}\left(\left(\frac{X-y}{X-m_k}\right)^{1+\rho} - 1\right)\right]$$

where γ measures the marginal utility of own income, β_i captures the overall importance of fairness relative to material payoffs, and ρ governs the curvature of the fairness loss. Higher ρ increases local curvature at the fairness benchmark and steepens tail penalties, reducing tolerance to both small and extreme deviations. For $\rho > 0$ the loss is strictly convex, with curvature increasing in ρ so marginal disutility rises more steeply as deviations grow. Whereas for a lower ρ the curvature increase, implying sharper penalties for large departures from m_k.

The first term, γy, is direct utility from own income. The second term is a CES loss centered at the fairness benchmark $m_k(a, q)$. It aggregates two symmetric components: a cost for overshooting the benchmark and a cost for undershooting it. The curvature parameter ρ controls how sharply the loss rises with proportional deviations: equivalently, the elasticity of substitution between these two sides of the deviation is $\sigma = 1/(1+\rho)$. Throughout, we take $\rho > 0$, which makes the loss strictly convex so that marginal disutility increases with distance from m_k.

Differentiating with respect to y yields

$$\frac{\partial V_i}{\partial y} = \gamma - \frac{\beta_i}{X(a,q)}\left[\left(\frac{y}{m_k}\right)^{\rho} - \left(\frac{X-y}{X-m_k}\right)^{\rho}\right],$$

which implicitly defines the optimal proposal y_i^* as a function of ρ.

In the quadratic model, small and large deviations scale in a mechanically predetermined way across all allocation sizes. The CES form relaxes this assumption by letting the responsiveness of marginal disutility vary smoothly with the curvature parameter ρ. The objective is not to redefine the fairness ideals themselves (SE, LE, L), but to test whether the evidence for pluralism in fairness ideals is robust to an alternative mapping from deviations to utility, that is, whether the estimated type shares λ_k are sensitive to the functional form of utility (or rather disutility). The parameter β_i continues to scale the relative weight placed on fairness, while ρ determines how the intensity of that concern changes with the size of the deviation.

The CES functional form, introduced to describe production functions by Arrow and Solow (1961), captures how easily one input can substitute for another through a single elasticity parameter. The curvature parameter sets how easily the two sides of the deviation trade places. If the elasticity is low, the loss acts like a maximum: the bigger deviation sets the value. If the elasticity is high, the loss acts like a weighted sum: the model swaps one violation for another with little cost. The familiar CES shapes appear in the iso-loss lines. At elasticity equal to one the lines form Cobb-Douglas curves. At elasticity equal to zero they form Leontief corners. When elasticity grows very large the lines turn almost straight, making the loss flatter and more forgiving. The idea still works like before, but now the data can tell us if people get more or less sensitive to unfairness as it grows, instead of assuming it's constant.

3.4 Structural model and estimation

I re-estimate the structural model on the original dataset from The Pluralism of Fairness Ideals. The data and replication files (written in Stata, C, and Pascal) were obtained from the authors' openICPSR repository, a platform for registering behavioural economic research for the sake of transparency and reproducibility. Because the original estimation routines were written in legacy languages and dependent on non-portable libraries, I first reconstructed the estimation environment in Python, replicating the authors' maximum-likelihood method, following as closely as possible the procedure in Cappelen et al. (2007). Second, I extend the estimation strategy by using an alternative efficiency loss specification.[1]

The general idea is that every person fits one of three latent ideals k, with probability λ_k. Each ideal has a share between $1 \geq \lambda_k \geq 0$, and all participants are assumed to belong to one of the three ideals, so that the sum of the shares equals one, $\sum_k \lambda_k = 1$.

For each type k choices come from a model that adds shocks to a type specific utility. People of the same type share one utility and one shock draw, people of different types differ only in their ideals. Within a type each subject keeps the same form, apart from fairness weight β_i that can vary between person. For subject i, the experiment yields $t = 1, \ldots, T$ allocation decisions. The reason for multiple allocation decisions is that some scenarios decision are compatible with multiple of the ideals (for ex a SE and a L would make the same decision), therefore by design another allocation decision is presented. Let

$$p_k(y_{it} \mid X_{it}, \beta_i)$$

denote the logit probability of the observed choice if subject i were of type k.

3.4.1 Estimation by Maximum Likelihood

I start with following the original papers method and estimate the models using maximum likelihood with a standard logit structure. Individual fairness sensitivities β_i are assumed to follow a lognormal distribution, with mean ζ and standard deviation σ. For each fairness type k, the model assign a likelihood to the sequence of observed choices $\{y_{it}\}_t$ which form

[1] For routine technical tasks such as debugging Python code and generating LaTex, I used OpenAI ChatGPT (GPT-5.1). The tool was not used to generate substantive text or theoretical arguments. All modelling decisions, interpretations, and the written content of the thesis are my own, with input from my supervisor.

the basis for estimation:

$$L_{ik} = \int \prod_{t=1}^{T} p_k(y_{it} \mid X_{it}, \beta)\, f(\beta; \zeta, \sigma)\, d\beta,$$

and the mixture likelihood over all types is

$$L_i = \sum_{k \in \{SE, LE, L\}} \lambda_k L_{ik}.$$

The structural parameters $(\lambda_k, \gamma, \zeta, \sigma)$ are obtained by maximising the summed log-likelihood $\sum_i \log L_i$.

3.4.2 Estimating via Efficiency Losses

To complement the MLE analysis, I also evaluate fit using a scale-free α-efficiency index to avoid fitting in cardinal utility units (utilities carry only ordinal information). For each observation (i, t) and ideal k, define $\alpha_{it} \in [0, 1]$ as the smallest fraction by which (m_k, X_{it}) can be radially contracted while the model's best allocation still reproduces the observed utility:

$$U(y_{it}; m_k, X_{it}, \theta, \rho) \;=\; \max_{y} \; U(y;\; \alpha_{it} m_k,\; \alpha_{it} X_{it},\; \theta,\; \rho)$$

Given ρ, I estimate $\theta = (\gamma, \beta)$ by minimising the average squared efficiency loss,

$$\min_{\theta} \; \sum_i \sum_t (1 - \alpha_{it})^2,$$

I assign each subject to the ideal k that gives the smallest mean efficiency loss, the resulting proportions λ_k match the MLE shares in the quadratic model. The criterion measures mistakes in a dimension-free efficiency scale, not in utility units, which makes it less sensitive to the curvature/scale of the CES penalty. When comparing ideals, ρ is held fixed: Table 4.3 then shows how classifications move across a grid of ρ values. In the reported CES runs I normalise γ (set to 25) and do not integrate over a lognormal β_i, hence ζ, σ are not reported in Table 4.1.

This criterion measures prediction errors in a unit-free efficiency space rather than in utility space, making it less sensitive to the curvature and scale of the CES penalty. It thereby provides a stable and behaviourally interpretable fit even when the log likelihood surface is

bumpy.

Both estimation methods share the same data, the same three fairness ideals, and the same underlying utility structure, only the estimation criterion differs. The MLE route sets the choice probabilities to match the data. The efficiency route lines up the utilities that are observed with the utilities that are forecast but also it does so through an index that does not change when the scale changes.

4 Results and Discussion

This chapter presents and discusses the results from re-estimating the pluralism model under a CES loss. First, I report how participants are allocated across the three fairness ideals (strict egalitarian, liberal egalitarian, libertarian) under the CES fit and compare the implied mixture shares with those from the quadratic model side by side, with both estimation methods. Second, as a robustness check, I vary ρ over a pre-set grid and track how assignments change: the goal is to separate identification of ideals from curvature and to highlight where classifications are stable versus fragile. We organise the evidence accordingly: Table 4.1 and 4.2 summarise the distribution and count of participants across ideals, and Table 4.3 shows the sensitivity of classifications to ρ. Taken together, these outputs assess whether the CES form better accommodates heterogeneity in fairness concerns without reinterpreting the underlying ideals, and whether that added flexibility materially changes how observed choices should be read.

Table 4.1: Estimates of the Structural Model

	Original Quadratic Model (Cappelen, 2007)		**Alternative CES Model**	
	(1) MLE	(2) α-efficiency	(3) MLE	(4) α-efficiency
λ^{SE}, share strict egalitarian	0.435	0.583	0.454	0.443
λ^{LE}, share liberal egalitarian	0.381	0.218	0.339	0.363
λ^{L}, share libertarian	0.184	0.198	0.206	0.193
γ, marginal utility of money	28.359	16.437	13.526	25.000
Log likelihood	−337.584	–	−369.557	–
Objective ($\sum(1-\alpha)^2$)	–	1.1429	–	64.631

Columns (1)–(2) report estimates from the quadratic-loss specification (Cappelen et al., 2007), while columns (3)–(4) report estimates from the CES-loss specification. Within each specification, odd columns use maximum likelihood estimation (MLE) and even columns use the α-efficiency.

Table 4.2: Distribution of Participants Across Fairness Ideals

Fairness Ideal	**Original Model**	**CES MLE**	**CES α**
Strict Egalitarian (SE)	42	44	43
Liberal Egalitarian (LE)	37	33	35
Libertarian (L)	17	19	18
Total	96	96	96

Counts indicate the number of participants classified under each fairness ideal based on model assignment. Both models classify the same total of 96 individuals, corresponding to the experimental sample.

Table 4.3: Sensitivity of Ideal Classification to the CES Curvature Parameter ρ

ρ	**Share of participants**						**Count of participants**					
	SE		LE		L		SE		LE		L	
	MLE	α	MLE	α	MLE	α	MLE	α	MLE	α	MLE	α
0.5	0.434	0.510	0.369	0.208	0.197	0.281	42	49	35	20	19	27
1.0	0.454	0.531	0.339	0.229	0.207	0.240	44	51	32	22	20	23
1.5	0.462	0.556	0.335	0.224	0.203	0.211	44	54	32	22	20	20
2.0	0.471	0.583	0.330	0.218	0.199	0.197	45	56	32	21	19	19

Each row corresponds to a different CES curvature parameter ρ. Shares sum to one across ideals for each ρ and correspond to the proportion of the 96 participants assigned to each type.

The first and most important thing to note is that the three fairness ideals remain robust under the CES model: each still claims a noticeable share, consistent with the idea that the underlying ideals are still present in the data. Although it is hard to be confident about the exact size of shares, the SE ideal is the largest under both measures and across all values of ρ. Using the α-efficiency objective, the lower the ρ, the more people are assigned to libertarians, by increasing ρ more people shift to be strict and LE. LE reacts most to curvature, its share grows when the loss function flattens and falls when the tails steepen. L moves in the opposite direction to LE. In contrast, the MLE mixture shares are more stable across ρ, with only minor reallocations across ideals as curvature varies.

4.1 Interpretation

Any interpretation of the precise shares of the ideals in the models should be taken with caution since the small sample and people at the threshold between ideals greatly affect the proportions of ideals. This is something that they note in the original article, the foremost conclusion is existence of ideals not the size of the shares.

One of the more interesting things to note is how we can explain why the share of libertarians can exceed that of liberal egalitarians in the CES model, especially for lower ρ? A simple way to see it is that when ρ is low the CES loss tolerates uneven splits. Choices that seem self motivated or close to a pure production based division are therefore more likely to be labelled libertarian instead of liberal egalitarian. As ρ increases, the CES loss becomes smoother and less kinked, approaching the quadratic model's circular shape.

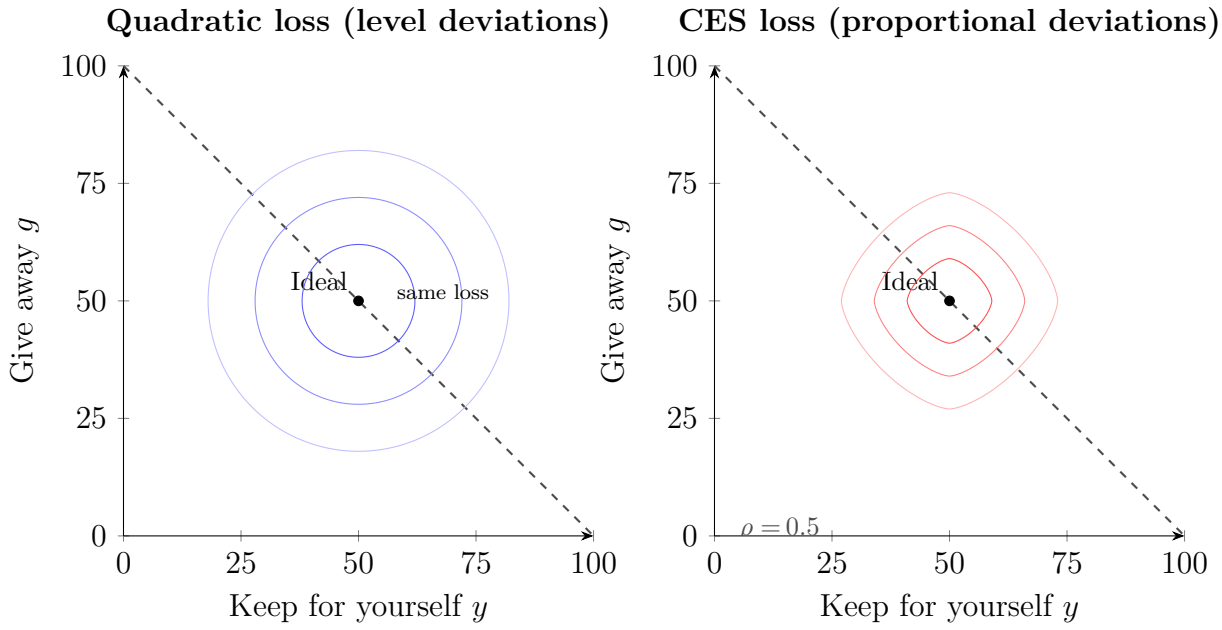


Figure 4.1: Iso-loss curves for Quadratic vs. CES loss ($\rho = 0.5$)

Figure 4.1 demonstrates one example of the difference between how quadratic and CES loss behave. On the left, we see that the quadratic form for equal unfairness forms circles around the ideal point. Any change that increases the level distance from the point increases the penalty. For example, multiplying every share by the same number while keeping the proportions fixed increases the unfairness score. Therefore, a change that gives every person more (a Pareto improvement) can sometimes be rated as a less desirable outcome than the initial allocation where everyone was worse off.

On the right side of the graph, iso-loss lines generated by the CES are plotted, which use relative gaps rather than absolute gaps. Multiplying every share by the same number leaves the unfairness score unchanged. Because the CES form keeps the penalty the same when only the scale changes, it satisfies the Pareto property.

4.2 Limitation

Several limits affect how the results should be read. The CES specification offers a different way to look at fairness ideals. The main finding in Cappelen et al. (2007), is that people hold different views, not the exact share of the population in each view. With only 96 subjects, the sample is too small to place reliable confidence intervals on the mixture weights or to claim precise sizes for the fairness groups. The reallocations that appear, most clearly the move toward strict egalitarianism, should be treated as an indicative result, not as a conclusion that we can put any statistical significance on.

Second, the sample does not allow for systematic exploration of heterogeneity within ideals. With only between 15–26 subjects per group, depending on the model, the data is insufficient to examine whether SE, LE, or L types differ along background variables such as age or gender. Any such comparisons would be underpowered and potentially misleading.

Thirdly, there's a trade-off between flexibility and parsimony, the CES model introduces another parameter, ρ, and therefore moves in the direction of a model that gains in adaptability but at the cost of simplicity. This increases the risk of overfitting the model to behavioural noise, instead of revealing further normative structure. The limitation is therefore a trade-off, the CES form may describe observed choices better, but it does so by adding an extra degree of freedom that can reallocate residual variation across types. This makes it important to weigh fit gains against complexity, for example, using penalised criteria.

5 Conclusion

This thesis re-estimate the pluralism model of Cappellen et al. (2007) with a CES loss function to see whether a curved fairness concern changes the way people are sorted into strict egalitarian, liberal egalitarian or libertarian groups. The main results are plain: the mix of ideals stays intact, the CES form pushes a little more weight toward strict egalitarianism, strict egalitarianism still dominates for almost every tested degree of curvature.

This thesis turns a hidden modelling choice into an open one. When the quadratic loss is dropped and curvature is allowed to vary, the earlier claim that pluralism exists does not depend on the exact modelling decision made. The CES model, therefore, acts as a robustness test for the central finding in the fairness literature: different normative ideals live side by side. This supports the view that the observed pluralism reflects genuine differences in fairness views rather than a result of a particular functional form.

Finally, although this thesis or the original article is not primarily policy-oriented, the results speak to the importance of recognising normative heterogeneity when thinking about redistribution or when fairness views are read in real-world settings.

References

Andreoni, James and John H. Miller (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism". In: *Econometrica* 70.2, pp. 737–753.

Arrow, Kenneth J. and Robert M. Solow (1961). *Contribution to the Theory of Growth*. Vol. 75. 1, pp. 65–94.

Bentham, Jeremy (1789). *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Bolton, Gary E. and Axel Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition". In: *American Economic Review* 90.1, pp. 166–193.

Cappelen, Alexander W. et al. (2007). "The Pluralism of Fairness Ideals: An Experimental Approach". In: *American Economic Review* 97.3, pp. 818–827.

Charness, Gary and Matthew Rabin (2002). "Understanding Social Preferences with Simple Tests". In: *Quarterly Journal of Economics* 117.3, pp. 817–869.

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren (2002). "Hardnose the Dictator". In: *American Economic Review* 92.4, pp. 1218–1221.

Fehr, Ernst and Gary Charness (2024). "Social Preferences: Fundamental Characteristics and Economic Consequences". In: *Journal of Economic Literature* 63.2, pp. 440–514. DOI: 10.1257/jel.20241391.

Fehr, Ernst and Klaus M. Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation". In: *Quarterly Journal of Economics* 114.3, pp. 817–868.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits (2007). "Individual Preferences for Giving". In: *American Economic Review* 97.5, pp. 1858–1876.

Forsythe, Robert et al. (1994). "Fairness in Simple Bargaining Experiments". In: *Games and Economic Behavior* 6.3, pp. 347–369.

Frohlich, Norman, Joe A. Oppenheimer, and Anja Kurki (2004). "Modeling Other-Regarding Preferences and an Experimental Test". In: *Public Choice* 119.1–2, pp. 91–117.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982). "An Experimental Analysis of Ultimatum Bargaining". In: *Journal of Economic Behavior and Organization* 3.4, pp. 367–388.

Hoffman, Elizabeth et al. (1996). "Social Distance and Other-Regarding Behavior in Dictator Games". In: *American Economic Review* 86.3, pp. 653–660.

Kagel, John H. and Alvin E. Roth, eds. (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Konow, James (2000). "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions". In: *American Economic Review* 90.4, pp. 1072–1091.

Niehans, Jürg (1990). *A History of Economic Theory: Classic Contributions, 1720–1980*. Baltimore: Johns Hopkins University Press.

Pareto, Vilfredo (1906). *Manual of Political Economy*. Milan: Società Editrice Libraria.

Rawls, John (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Samuelson, Paul A. (1947). *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press.

6 Appendix

Directory of Aids

AI-appendix

The following aids were used in the preparation of this thesis. All substantive content, modelling decisions, interpretations, and written text are the independent work of the author. No AI-generated content has been inserted into the thesis.

Artificial Intelligence Tools

- OpenAI ChatGPT (GPT-5.1, ChatGPT-5 Thinking mode)
  Used exclusively as an aid for:
  - generating and debugging Python/Stata code,
  - generating LaTex snippets (tables, equations, figures, formatting),
  - clarifying econometric and behavioural-economic concepts,
  - idea generation during the research process.

  No AI-generated content has been directly inserted into the thesis.
- Grammarly
  Used for spelling, grammar, and linguistic surface revision.

Insights gained and quality improvement

AI helped with tasks such as translating the legacy code from the original article into modern executable Python code, a routine task that would have taken a lot of time to do manually, and without any real impact on the substance of the thesis. This freed up time for doing actually quality improving work, such as reading more literature, developing the models, and testing alternative specification of the model that would otherwise have been difficult to consider. It also assisted with practical tasks like extracting large pdfs of content that is relevant to the research questions and generation of LaTex expressions. They did not replace the need for me to think through concepts, choose models or explain what the results mean.

Risks and mitigation

It is a well-known fact that AI models are prone to hallucinations, increasing the risk of hard to notice errors. Every contribution, of code or LaTex equation that the AI produced was therefore checked carefully. This ensured that the AI sped up the work up without lowering the correctness or the dependability of the results.