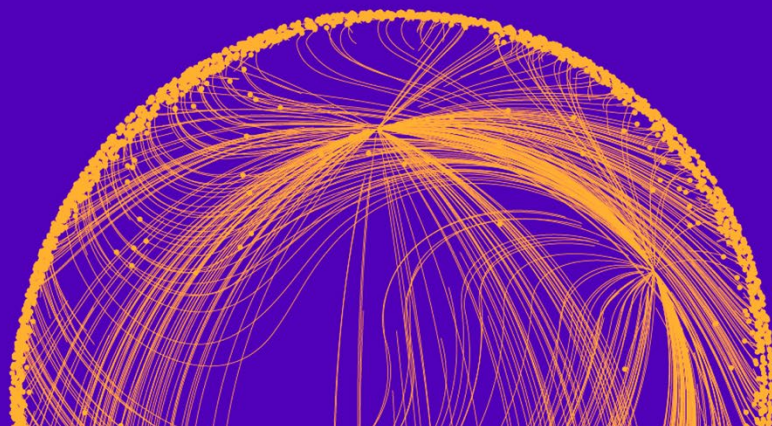


HELSINKI GSE DISCUSSION PAPERS 14 · 2023

Personal and societal conflict of distributive principles and preferences

Ola Andersson
Marco Lambrecht
Topi Miettinen



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



A! Aalto University

Helsinki GSE Discussion Papers

Helsinki GSE Discussion Papers 14 · 2023

Ola Andersson, Marco Lambrecht, and Topi Miettinen:
Personal and societal conflict of distributive principles and preferences

ISBN 978-952-7543-13-9 (PDF)

ISSN 2954-1492

Helsinki GSE Discussion Papers:
<https://www.helsinkigse.fi/discussion-papers>

Helsinki Graduate School of Economics
PO BOX 21210
FI-00076 AALTO
FINLAND

Helsinki, August 2023

Personal and societal conflict of distributive principles and preferences

Ola Andersson* Marco Lambrecht † Topi Miettinen‡

August 4, 2023

Abstract

Distributive justice principles are commonly used in societal debates to support policy positions and argue for their legitimacy. Prominent formal incarnations of such principles are the Rawlsian, Utilitarian, Efficient transfer (e.g. Pigou-Dalton), and Leaky-bucket transfer (e.g. Hammond) principles, but could also be formulated by the minimization of the Gini coefficient, for instance. We elicit preferences for these principles and collect distributive choices over income distributions in a large online experiment with participants from all walks of life. We randomly assign half of the sample to a resolution treatment when preferred principles are inconsistent with choices and study how personal conflicts are resolved. Except for the Gini principle, we find no statistically significant difference across initial subscriptions. However, among participants who are given the opportunity to resolve personal conflicts, the majority abandon their subscription to the principle. Efficient transfer principle is an exception, showing significantly more subscriptions than other principles after resolution. Additionally, we survey preferences over redistributive policies and find that the resolution mechanism reduces societal conflict. This finding has important implications for the elicitation of policy preferences in applied work.

JEL classification: C91, D63, D71, D91

Keywords: social choice, Socratic method, policy preferences

*Uppsala University, UCFS & Research Institute of Industrial Economics (IFN); email: ola.andersson@nek.uu.se

†Helsinki GSE & Hanken School of Economics; email: marco.lambrechth@hanken.fi

‡Helsinki GSE & Hanken School of Economics; email: topi.miettinen@hanken.fi

We thank Alain Cohn, Ben Grodeck, Kaisa Herne, Peter Matthews, Kirby Nielsen, Stefan Traub, and the audiences at Helsinki GSE, Uppsala University, TSE, ESA Bologna 2022, ESA Lyon 2023, and the 43rd Annual meeting of Finnish economists for their constructive comments and suggestions. Funding from the Yrjö Jahnsson Foundation (grant no. 20177011) and the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1 Introduction

Individual preferences over distributive policy and income distributions vary substantially by gender, income position, political orientation, beliefs about distribution, immigration, and institutions, and by other-regarding attitudes (Cohn et al., 2022; Karadja et al., 2017; Almås et al., 2020; Deffains et al., 2016; Kuziemko et al., 2015; Alesina et al., 2023; Fehr et al., 2021). Recently, it has been argued that polarization of opinions across the political party platforms is on the rise (Fiorina et al., 2008; Abramowitz and Saunders, 2008). The advancement of affective polarization (negative affect towards supporters of the opposing political party) observed in the U.S. also extends to other countries (Boxell et al. 2022), and may have been associated with increased polarization of policy preferences. Moreover, there is evidence that polarization affects the quality and size of political institutions (Lindqvist and Östling, 2010; Hetherington and Rudolph, 2015) and satisfaction with democracy (Hoerner and Hobolt, 2020). Some argue that political views are, to an increasing extent, affective and expressive rather than rational and consistently focused on facts and policy positions (Abramowitz and Webster, 2016; Robbett and Matthews, 2018). Inconsistencies in judgment and choice pose a challenge to normative analysis in general (Bernheim, 2009, 2016), and to distributive policy and social choice in particular (Fleurbaey and Schokkaert, 2013) as the traditional welfare analysis in economics builds on individual revealed preference and seeks policy improvements using the revealed preference profile as the basis for welfare evaluation.

In this paper, we investigate people’s preferences over distributive principles and their consistency with actual choices over distributive outcomes. And, if revealed principles and choices are inconsistent, we examine how people resolve these inconsistencies by making them salient to a randomized sample. We study if these resolutions causally affect distributive policy preferences at a personal level. Finally, we study whether these effects are systematic and causally affect societal polarization in views over distributive principles and policies as a consequence.

Consistent logical arguments supporting policies and ideals are seen as a strength in societal debate.¹ Such debate may also provide opportunities for pointing out inconsistencies in individual views and arguments. The idea that debate could reveal inconsistencies in one’s thinking features prominently in the *Socratic method* illustrated by Plato in his early writings, and John Stuart Mill advanced this idea in societal decision making on the grounds of opening up for the opportunity for each decision maker or politician to correct their initially erratic judgment (Elster, 1998, p.4). Relatedly, the idea that rational deliberation, in groups

¹Falk and Zimmermann (2017); Schwardmann et al. (2022) show that consistent thinking and logical argumentation are valued by others and promote success in an experimental labor market and in debating competitions, respectively. Schwardmann et al. (2022) moreover show that presenting logical arguments causally changes the preferences and views of the presenter in favor of these positions, too.

or individually, leads to greater societal agreement is clearly present in the contractarian tradition (Locke, Hobbes, Rousseau, Kant, Harsanyi, Buchanan, Rawls, Nozick, Gauthier, among others), which grants normative legitimacy to particular social institutions if they can be conceived as an outcome of rational negotiations or deliberation by members of society under some ideal impartial circumstances, a.k.a. original position.² Indeed, Rawls (1971, p.17) states that “[u]nderstood in this way, the question of justification [of fairness] is settled by working out a problem of deliberation: we have to ascertain which principles it would be rational to adopt given the contractual situation. This connects the theory of justice with the theory of rational choice.” By resorting to the idea of deliberation in an original position behind the veil of ignorance, he concluded with the *maximin* principle (also called *difference* principle), which requires that the well-being of the worst-off individuals in society should be maximized.³ Similar arguments to justify *Utilitarianism*, i.e. maximization of a weighted sum of utilities, have been put forth by Harsanyi (1953, 1955, 1977). Two prioritarian rules, *Efficient transfer*, which is closely related to the Pigou-Dalton transfer principle Dalton (1920), and *Leaky-bucket transfer* principle, closely related to the Hammond transfer principle Hammond (1976), call for redistribution from richer to poorer, alas, the first one approves such changes only if no resources are lost in the transfer. From an academic perspective, distributive social choice principles provide means for expressing ideal characteristics of income distributions and redistribution clearly and non-ambiguously. They establish normative benchmarks for evaluating the performance of (re)allocation rules, and reflect fundamental properties of what is meant by fairness, efficiency, and equal distribution.

The idea that deliberation allows individuals to revise their positions to form consistent views is reflected in Gilboa and Schmeidler (2001) and Gilboa (2010). They advance a role of normative, i.e. prescriptive, theory in consulting people to make good decisions. They propose using an approach closely related to the Socratic method: make people aware when violating a principle, and consult whether they would like to change their decisions to align them with the principle. Nielsen and Rehbeck (2022) apply this idea to decision-making in the laboratory and show, for all individual rational choice axioms (including axioms for decisions under risk), that study participants revise their choices to align with the principle. Thus, they find strong descriptive evidence supporting the proposed role of the prescriptive rational theory.

²Recently, Sugden (2018) advanced the contractarian approach to behavioral welfare economics. There is also a natural rather than normative theory of justice which promotes the view that adaptive forces might favor particular fairness views and principles when societal debate (or gossiping) shapes the fairness norms in societies (Binmore et al., 1994; Binmore, 1998).

³This is the arguably most prominent social choice principle, particularly in the Anglo-Sax world where “Rawls is widely considered as the most important political philosopher of the 20th century” (Encyclopaedia Britannica, 2023).

In this study, we apply the methods of Nielsen and Rehbeck (2022) to the distributive principles of social choice theory in a pool of 2,295 British citizens from all walks of life. We randomly assign a distributive justice principle (in our case either the *Rawlsian*, *Utilitarian*, *Efficient transfer*, *Leaky-bucket transfer*, or the minimization of the *Gini* coefficient) and ask participants whether they would like that distributive choices comply with that principle. We then let them choose earnings distributions for anonymous participants in a different location, thus acting as social planners or impartial spectators. Subsequently, we conduct an experiment by applying the Socratic method to half our sample: if the distributive choices violate the principle they subscribed to, we ask half of those whether they would like to change their distributive choice, unsubscribe the principle, or both, or neither. We call this the Socratic treatment. All choices are incentivized. Finally, we survey preferences over distributive policies to study the causal effect of the Socratic method on policy preferences.

Using this setup, we study the following set of questions. Which of the principles tested here, i.e., *Rawlsian*, *Utilitarian*, *Efficient transfer*, *Leaky-bucket transfer* or *Gini*, do people subscribe to (revealed preferences over principles)? How do people distribute income to a group of strangers in a position of a social planner (revealed distributional preferences)? How do people react to conflicts between their revealed preferences over principles and their revealed distributional preferences when confronted with those (Socratic method)? Moreover, we extend the experimental setup to allow us to study the causal effects of applying the Socratic method. We ask: Does the Socratic method affect (i) what principles of distributive justice people prefer? More practically, (ii) does it affect preferences over societal policies related to redistribution (income, wealth, inheritance taxation, basic income, the funding of primary education, and so forth)? Specifically, we are interested in whether the Socratic method might have a causal effect on societal disagreement, i.e. variance, regarding the support for each of the (i) principles and (ii) each of the policies, thus reducing the polarization of opinions.

Nielsen and Rehbeck (2022) observe quite high subscription rates to individual decision-making axioms, sometimes called rationality principles. However, it is worth noting that the social choice principles of our study differ substantially from rationality principles, as they in contrast to the latter may be mutually exclusive. Indeed, it is a matter of preference whether one wants to subscribe to the Rawlsian or the Utilitarian principle, for instance. Clearly, the distributions consistent with the two are often quite different.⁴ Thus, our principles may contradict each other in ranking two given income distributions and, therefore, consistent decision makers in our study should not subscribe to all principles. In other words, while

⁴In fact, these two constitute the two opposite extremes of a preference scale where the curvature of the utility in a social welfare function is varied (Atkinson et al., 1970; Johansson-Stenman et al., 2002; Fisman et al., 2017).

rationality axioms may be complementary (and when applied jointly, lead to rational decision makers as defined in the von Neumann Morgenstern expected utility model), distributive justice principles should, in many cases, be considered as substitutes. Thus, we expect much lower subscription rates in our experiment.

We find that the majority of individuals who initially subscribe to a particular distributive principle revise the subscription when exposed to the Socratic method. The Efficient transfer principle (any opportunity to pass money from a higher to lower income individual should be used if the transfer does not undermine efficiency) is the only distributive principle that survives the Socratic method. Only a minority of Efficient transfer subscribers revise their subscriptions. At the same time, Efficient transfer has the highest rate of consistency with distributional choices – both among subscribers and non-subscribers. Yet, there is also a fraction of subscribers who do revise their Efficient transfer subscriptions. Interpreting the evidence using the approach of Nielsen and Rehbeck (2022), our decision makers consider their subscriptions to principles as mistakes, rather than their distributive choices. The only exception is the Efficient transfer principle, for which subscriptions are quite robust. To sum up, measuring distributive preferences solely by means of eliciting preferences over principles might lead to a biased and not fully comprehensive picture of what people really prefer. Then again, a substantial fraction of participants consider their choices mistakes, especially when those choices conflict with the Efficient transfer principle. Thus, eliciting distributive preferences through choices without allowing any reconsideration against one’s principles might generate a biased and not fully comprehensive measure of distributive preferences either. Resolving personal conflicts might provide a more robust and comprehensive, albeit time-consuming and demanding, elicitation method.

Let us then turn to distributive policy preferences and their connections to distributive principles. Previous literature has shown that views regarding policies vary according to age, income level, gender, and social preferences. Such variation is also visible in our data. However, does the Socratic method have any meaningful effect on the variation of individual preferences over societal policy? We find that applying the Socratic method, in some cases, causally decreases the polarization of expressed policy support. For example, there is less variation in the support for redistribution among those treated with the Socratic method than in the control group. The strongest effects are visible among individuals who receive Socratic treatment for their subscription to the Rawlsian principle: in that case, the variation in support for redistribution, basic income, inheritance tax, and capital gains tax is significantly reduced. Thus, the Socratic method reduces societal disagreement regarding key policies significantly. Indeed, the Socratic method typically shifts individual opinions away from the extremes towards the average opinion in society. We see similar effects of reduced disagreement in the subscriptions to the distributive principles.

In the descriptive social choice literature (Gaertner and Schokkaert, 2012; Schokkaert and Tarroux, 2021), among the closest studies are those reported in Amiel and Cowell (1999) who investigate distributive preferences from a descriptive perspective by eliciting preferences by what they called verbal and numeric methods. These two come close to what we call subscriptions to principles and distributive choices in Stages 1 and 2 of our experiment, respectively. Like us, Amiel and Cowell elicit preferences over principles and preferences over distributive choices from each participant. Also like us, they allow for revising the expressed preference for a principle after making a distributive choice between a pair of distributions. Yet, they do not reveal inconsistencies between the two; thus, the Socratic method, as defined above, is not applied. Moreover, there is no overlap between the distributive principles studied here and those studied by Amiel and Cowell.

Our study also relates to the paper by Pirttilä and Uusitalo (2010), which contrasts two survey methods to elicit personal efficiency-equity trade-offs in a representative sample of the Finnish population. First, they find considerable personal conflicts between the two elicitation methods. Second, these conflicts are significantly correlated with own position in society. Like Amiel and Cowell, they do not reveal any information about conflict nor elicit any decision on how to resolve that conflict. In a recent paper, Schönegger and Grodeck (2022) apply the methodology of Nielsen and Rehbeck (2022) to study the method of narrow reflective equilibrium in population ethics. They provide experimental data on conflicts between intuitions about general moral principles and hypothetical concrete cases. In their setup, participants face alternatives that differ with respect to population size as well as the levels of happiness of citizens. Due to their focus on population ethics, where they vary the size of the population, they do not study the same principles as we do. Moreover, in contrast to us, they use a status-quo design where participants are asked if they want to keep the hypothetical status quo or transition to an alternative society. In contrast, we ask subjects to pick their preferred allocation of money to ten individuals out of two proposals.⁵ Like us, they find that decision makers are more likely to revoke their commitment to abstract principles than to revise their choices in concrete cases. The exceptional nature and robustness of the Efficient transfer principle, as opposed to other principles, is the key distinctive result of our study not featured in theirs. None of the above-mentioned studies uses the Socratic method as a treatment to study how the deliberative process causally affects distributive justice and policy preferences and their polarization.

Ever since Frohlich et al. (1987), the social choice principles studied here have been examined in questionnaire, hypothetical, and incentivized decision experiments. However, none

⁵To see how this affects conclusions, note that when they for instance study the Pareto principle, they challenge the principle by introducing a status quo and offering an alternative where one population is made a lot better off, with the other population only being made slightly worse off. In our framework, without a status quo, this would not pass as a conflict between the Pareto principle and choice.

of the existing studies have examined the robustness to the Socratic method nor the causal effect of the latter on preferences over societal policy. Among others, Amiel et al. (2009), find descriptive support for the Pigou-Dalton transfer principle across countries (see also Gaertner and Schokkaert (2012)). In individual decision-making tasks, Frohlich and Oppenheimer (1992) and later Traub et al. (2005) find little support for the Rawlsian maximin or leximin nor the minimization of Gini but rather combinations of Rawlsian and Utilitarian principles turn out popular.

A recent literature strives to understand how redistributive-policy preferences are causally affected by information provision. Thereby, these studies focus on the causal effects of revealing information about inconsistent beliefs, whereas our focus is on revealing information about inconsistencies in preferences. Cruces et al. (2013) and Kuziemko et al. (2015) reveal information to subjects about the true distribution of wealth and study its causal effect on policy preferences. Whereas Cruces et al. (2013) find a causal effect of information on redistributive preferences, in Kuziemko et al. (2015) the effect on policy preferences is minor despite the large correction in beliefs about inequality. Contrary to our study, the effect on societal disagreement is not studied. Similar information treatment protocol experiments and their effects on a particular policy preference have been implemented by Karadja et al. (2017) as well as Fehr et al. (2022) on redistribution and Bastani and Waldenström (2021) on inheritance tax. Dunaiski and Tukiainen (2023) exploit a naturally occurring information treatment to study effects on redistributive preferences. We find that the Socratic method causally reduces polarization in both distributive and societal policy preferences, which has implications for the elicitation of policy preferences.

There is an emerging experimental literature on group deliberation in political science. Grönlund and Herne (2022) study the effects of small-group deliberation, mediated or not, on various outcomes of interest, such as distribution, individual views, and polarization of individual views. Strandberg et al. (2019), for instance, find that deliberative norms causally reduce societal polarization of opinion due to deliberation in like-minded groups. Ferreira et al. (2023) find that group deliberation reduces within-group polarization of opinions regarding redistribution, and the prime driver of this effect is group identity. We contribute to this literature by showing that individual deliberation, due to the application of the Socratic method, also reduces polarization (among the participants in the experiment), in a setting where group identity and social norms are mute.

The rest of the paper is organized as follows. Section 2 summarizes the experimental design and the protocol. Section 3 presents the results. In section 4, we discuss the implications of our findings in a broader context, and section 5 concludes.

2 Experimental Design

In our experiment, each decision maker is randomly assigned to one of the following social choice principles: *Rawlsian*, *Utilitarian*, *Efficient transfer*, *Leaky-bucket*, or *Gini-coefficient minimization*. Subsequently, each decision maker is presented with that principle and its polar opposite, which we call the *anti-principle* in what follows. The anti-principle is introduced to control for experimenter demand effects.⁶ The decision maker then decides whether to subscribe to each of these: subscription to both, only one, or neither is allowed (see *Subscriptions to principles* in subsection 2.2.1 below).

Subsequent to the subscription decisions, the decision maker goes through a series of eight distributive decision tasks (see *Distributive choice* in subsection 2.2.2 below). In each decision, there are two income distributions. The decision maker decides in the role of an impartial *quasi-spectator* (Konow, 2009), who has no own monetary stake in the decision, which of these two distributions to choose. The tasks are uniquely tailored to the principles assigned so that each choice of a distribution will violate either the principle or the anti-principle.

If one or more of the decision maker’s distributive choices violate a principle subscribed to, we call this a *personal conflict*. One of the key novelties of our design is that we randomly assign decision makers to the Socratic method (Socratic treatment) in which they will learn about personal conflicts if there are any, and be offered an opportunity to revise their subscription to the (i) principle, (ii) distributive choice, (iii) both, or (iv) neither (see *Socratic stage* in subsection 2.2.3 below). The other participants are not confronted with their potential conflicts (Baseline treatment).⁷

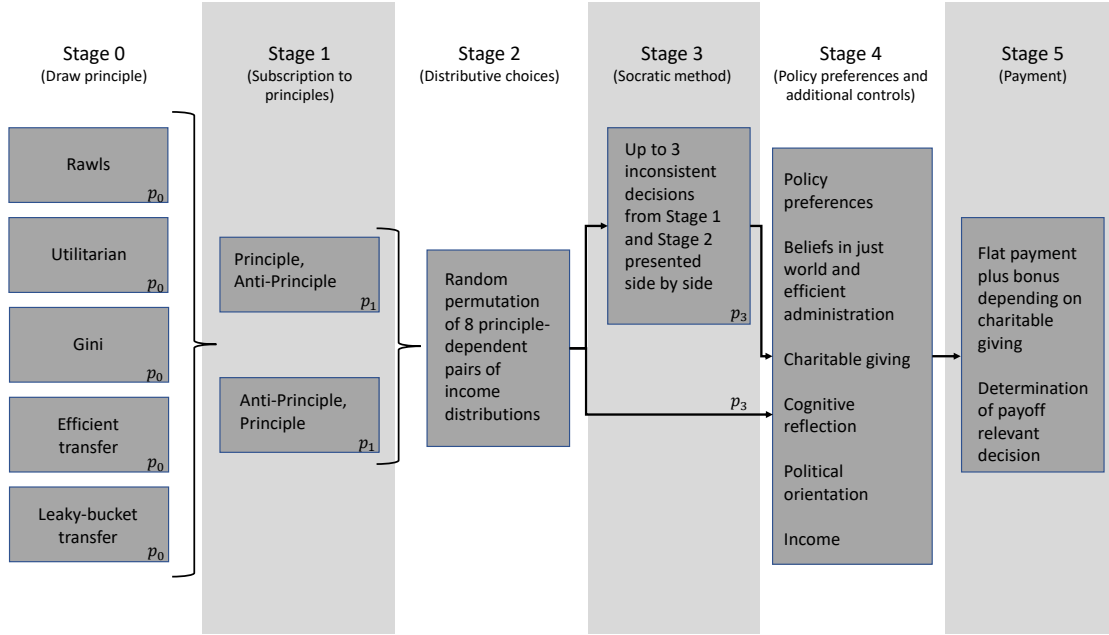
The experimental sessions conclude with the elicitation of policy preferences and additional control variables for our analysis. The design was preregistered and received ethical approval.⁸ Overall, we employ a 5×2 factorial design to examine, for each of the five principles, the effects of the Socratic treatment on redistributive preferences and societal (dis)agreement. We provide a flowchart of the randomization procedures and the five different stages of the experiment in Figure 1, and provide screenshots of the first four stages in Appendix B.

⁶This approach is similar to that of Nielsen and Rehbeck (2022). In our design, each decision maker may only subscribe to one principle and the corresponding anti-principle. Since they are the polar opposite of each other, each choice can only violate one principle at a time. This facilitates the design of the resolution of conflicts and gives stronger control of the causal effects of the Socratic method.

⁷Causal effects of the Socratic method on any other preferences were not studied in Nielsen and Rehbeck (2022).

⁸For the preregistration, see *The Open Science Framework*, osf.io/ycd9f. IRB was granted by the *Research Ethics Committee* of the Hanken School of Economics.

Figure 1: Flowchart of experimental stages



Note: $p_0 = \frac{1}{5}$ is the probability of each of the principles being assigned to the participant. $p_1 = \frac{1}{2}$ is the probability of the order of principle and anti-principle being shown to the participant. Finally, $p_3 = \frac{1}{2}$ is the probability to be randomly assigned to the Socratic treatment.

2.1 Distributive principles

Our study includes five distributive principles: the Rawlsian, Utilitarian, Efficient transfer, and Leaky-bucket transfer principles and the minimization of the Gini coefficient. The Rawlsian and the Utilitarian principle compare the distributions directly. The Efficient transfer, (Dalton, 1920) and the Leaky-bucket transfer (Hammond, 1976) principle rather evaluate distributions based on whether they can be reached through implicit *transfers* that satisfy the principles. The Rawlsian and the Utilitarian principles are typically used in the literature to normatively characterize societal distributions of income, wealth, societal resources, utility, or even of rights and liberties (Rawls, 1971; Harsanyi, 1977; Frohlich and Oppenheimer, 1992). In that vein, the Rawlsian principle calls for maximizing (the utility of) the allocation to the least advantaged in the target population. In contrast, the Utilitarian principle calls for maximizing the weighted sum of (utilities generated by) the allocations.

In our experiment, each decision maker chooses between two income allocations to ten anonymous recipients. Thus, as in Frohlich et al. (1987); Herne and Suojanen (2004), what we label as the Utilitarian principle is more narrowly presented as the maximization of average or total earnings, i.e. an *efficiency* principle. We decided to go for this approach, as it is easier to present unambiguously to the decision makers. It is also unambiguous to identify a violation of efficiency in the participant's choice. Theoretically, the Utilitarian principle

reduces to the efficiency principle only if each member of the target population receives an equal weight, and individual utility is linear in income (Blackorby et al., 2002). Indeed, if individual utilities were of the CRRA form, then the Rawlsian and the efficiency principle would be at the opposite ends of a scale where the coefficient of relative risk aversion is varied (Atkinson et al., 1970; Carlsson et al., 2003; Fisman et al., 2017).⁹

Gini is not really a social choice principle. Still, we decided to include it as the Gini coefficient features prominently in the literature on wealth and income inequality, and it is widely used in applied policy work in national and international arenas. The Gini coefficient is based on the Lorenz curve and relates to the transfer principles (Sen, 1973). Despite the fact that understanding and calculating this measure is certainly not trivial, its popularity and relevance motivate a thorough examination.¹⁰

2.2 Principles, choices, Socratic method and policy preferences: the stages of the experiment

Each decision maker goes through up to five stages, depending on whether they subscribed to a principle and whether they are in the Baseline or Socratic treatment: (1) subscription to principle, (2) distributive choices, (3) Socratic method (only Socratic Treatment), (4) policy preferences and control variables, and (5) payment. We select one of the decisions from stages (1)-(3) to be payoff relevant for the recipients. Incentivized donation to charity in the fourth stage is always payoff-relevant. In what follows, we describe each stage in detail.

2.2.1 Stage 1: Subscriptions to principles

All decision makers begin the experiment with the principle subscription stage after being randomly assigned to one of our five principles.¹¹ They receive general instructions which explain the stage. In particular, they learn that decisions are inconsequential for the monetary payoff of our decision makers, but affect payments that ten strangers receive (none of which are decision makers in our experiment). Our decision makers learn that we will present them with a principle that can be used to identify a preferred of two given - but unknown - income

⁹We chose not to include principles that combine elements of the Rawlsian and the Utilitarian principles, such as the Utilitarian with a floor constraint (Frohlich and Oppenheimer, 1992; Traub et al., 2005) due to complexity issues or difficulties in defining them in a sufficiently general manner.

¹⁰Acknowledging its complexity, we have carefully evaluated different options of how to formulate an economic principle of Gini coefficient minimization. We refrained from providing mathematical details and formulas, instead opting for an intuitive description. For the exact implementation, see Appendix B.

¹¹We assign only a single principle and its anti-principle to each of our decision makers. Whereas we acknowledge that it would be interesting to know which set of principles people want to subscribe to, such design would complicate the analysis considerably because each distributive choice in Stage 2 could conflict with several principles at the same time, for instance.

distributions. They can then evaluate the principle and decide to subscribe to it, implying that it will automatically choose, in a randomly drawn pair of income distributions where the principle is relevant, which distribution to implement. Alternatively, the decision maker can choose not to subscribe, in which case the decision in the randomly drawn pair will not be made based on the principle, but on the individual decision of the decision maker.¹²

We control for experimenter demand by offering a chance to subscribe to an anti-principle in addition to the principle itself. This anti-principle states the polar opposite of the principle.¹³ This means that each decision maker has exactly two consecutive subscription decisions to make. They can subscribe to neither, only the principle, only the anti-principle, or both.¹⁴ The exact wording of the principles and anti-principles is as follows:

- (*anti-*)Rawls: Program A is preferable compared to program B if the lowest income in program A is larger (*smaller*) than the lowest income in program B.
- (*anti-*)Utilitarian: Program A is preferable compared to program B if the sum of all incomes of program A is larger (*smaller*) than the sum of all incomes of program B.
- (*anti-*)Gini: Consider two randomly chosen individuals in program A and two in program B. Program A is preferable compared to program B if the expected income difference between individuals (in proportion to the average income of the respective program) is smaller (*larger*) in program A than in B.
- (*anti-*)Efficient transfer: Program A is preferable compared to program B if there is a way to re-allocate income starting from program B and yielding program A such that re-allocations only involve transfers from individuals with higher (*lower*) incomes to individuals with lower (*higher*) incomes.
- (*anti-*)Leaky-bucket transfer: Program A is preferable compared to program B if there are two or more persons such that the poorest of them earns more (*less*) in A than in B, and the total earnings in A are at most (*least*) as much as in B. All other persons receive the same under both programs.

Notice that the Pigou-Dalton and Hammond principles, in their original definitions, evaluate single transfers from a status quo distribution. Our transfer principles are designed

¹²In practice, the decision in the randomly drawn pair is made based on the decision made in Stage 2. For details on the random draw to determine the relevant decision, see section 2.2.5.

¹³For example, if the principle identifies distribution A as better than distribution B, then the anti-principle identifies distribution B as better than distribution A. The principle and its anti-principle are presented on separate screens and in random order.

¹⁴However, we consider decision makers who subscribe to both as a distinct group which we exclude for parts of our analysis (653 decision makers in total). For details, see section 3.

to compare income programs for ten random individuals without a status quo (unlike in Schönegger and Grodeck (2022)). Therefore, our Efficient transfer and Leaky-bucket transfer principles differ slightly from the Pigou-Dalton and Hammond principles, respectively. In particular, the Efficient transfer principle calls for making any transfers from richer to poorer individuals when no income is lost in such transfers. This is equivalent to re-ordering individuals in both distributions from rich to poor, setting a status quo, and applying a sequence of rank-preserving Pigou-Dalton transfers. Moreover, the Leaky-bucket transfer principle calls for increased earnings for the poorest in a subgroup of individuals while the total earnings of this group are allowed to decrease. Such a change can be achieved by re-ordering the individuals in the group from rich to poor, setting a status quo, and implementing a sequence of rank-preserving Hammond transfers.¹⁵

In our study, arguably, some principles are more difficult to comprehend than others. Meanwhile, each principle and its corresponding anti-principle are of similar complexity. Differences in subscription rates between principle and anti-principle may thus serve (to some degree) as an indication of how well subjects understood the principles. The variation in complexity across principles is likely to affect subscription rates and has implications for later decisions in the experiment (in particular, during the revision stage). Indeed, this is part of the focus of the study.

2.2.2 Stage 2: Distributive choices

In the second stage of the experiment, each of our decision makers makes distributive choices as an impartial non-stakeholder (quasi-)spectator. The instructions explain that decisions are inconsequential for the payoff of our decision-makers, who receive a fixed fee, but instead potentially determine the payment for a group of ten anonymous recipients. The decision-makers are told that none of these participate in the current experiment.¹⁶ Moreover, the distributed incomes are expressed in an experimental currency unit without a known exchange rate, whereas our decision-makers are paid in British pounds. The magnitudes of the recipient payoffs are in thousands of currency units, i.e., numbers corresponding to levels of sterling-valued gross monthly salaries in UK.

In each decision task, decision makers choose between a pair of income distributions. They go through a series of eight binary distribution decisions tailored to each principle so that

¹⁵Clearly, the efficient transfer principle respects the efficiency principle. The Leaky-bucket transfer principle calls for making any transfers from richer to poorer irrespective of whether income is lost or not. It thus allows transfers that violate the efficiency principle, but it always promotes the Rawlsian leximin principle. Specifically, when the transfers are allocated to the lowest income earner, Leaky-bucket transfer promotes the ideal of the Rawlsian difference or maximin principle.

¹⁶These ten participants are in fact members of a convenience subject pool in a different country than UK.

one of the choices in each pair conforms with the principle and violates the anti-principle, whereas the opposite holds for the other choice. For each pair of income distributions, decision-makers must choose one and exactly one distribution. We do not explicitly mention that the principles of the previous stage are applicable to determine a favorable distribution in each pair. Our experimental algorithm compares the subscription decisions of the first stage to the distributional choices of the second stage and identifies any contradictions between the two (personal conflict). We designed our decision tasks to make the likelihood of personal conflicts high, building on the tensions created by each principle on the pair of distributions. To identify such pairs, we use techniques based on salience, skewness, common consequence, common ratio, and trade-offs in efficiency and equity, which are known to have led to violations in previous literature (Starmer, 2000; Gaertner and Schokkaert, 2012; Frydman and Mormann, 2018; Dertwinkel-Kalt and Köster, 2020; Schneider and Leland, 2021). For details on the distributional pairs see Appendix A.

2.2.3 Stage 3: Socratic method

In the third stage of the experiment, we apply the *Socratic method*. Half of the pool of decision-makers is randomly assigned to the Socratic treatment but only those with personal conflicts will participate in this stage. Decision-makers in the Socratic treatment without personal conflicts or decision-makers in the Baseline treatment skip this stage.

Decision-makers that enter this stage have thus committed to the principle (or anti-principle) but subsequently, in the second stage, made decisions violating this principle (or anti-principle). They now have to decide how to resolve the conflict. To keep the maximal length of the experiment similar across treatments, each decision-maker takes a resolution decision in at most three personal conflict situations. If they have more than three personal conflicts, decision makers get to resolve a random subset of three of their conflicts.

For each resolution decision, the decision-maker gets the opportunity to revise either of the decisions in the first and the second stage which led to the conflict. For each conflict, decision-makers are presented with the (main or the control) principle they subscribed to, as well as the pair of income distributions and the choice that contradicts the principle. They may freely choose to change their subscription to the principle, their choice between the two distributions, both, or neither.¹⁷

¹⁷As mentioned before, there are differences in the eight binary distribution decisions between assigned principles. These differences may influence the revision decisions in the third stage.

2.2.4 Stage 4: Policy preferences and additional controls

The fourth stage of our experiment is the same for all decision makers independent of the treatment. It elicits societal policy preferences and control variables featured in the literature on redistributive policy (Kuziemko et al., 2015; Karadja et al., 2017; Almås et al., 2020; Bastani and Waldenström, 2021; Cohn et al., 2022) relevant to our analysis. In particular, we ask participants whether they agree with the following statements:¹⁸

- *redistribution*: A society should aim to equalize incomes.
- *inheritance tax*: A society should have an inheritance tax.
- *capital gains tax*: A society should have a capital gains tax.
- *basic income*: A society should have an unconditional basic income paid to every individual.
- *early-life education*: It is more important to direct funds to primary education rather than to increase the benefits of the poorest 5% of the people in society (assuming the expenditures needed are the same in the two cases).
- *progressive income tax*: It is better to fund the key activities of the public sector by an inheritance tax than an increase in the progression of the income tax (assuming the effect on government revenue is the same in the two cases).

The last two statements were added to force participants to trade off competing goals. Additionally, we elicit beliefs in a just world, views on whether income tax affects effort, views about the efficiency of public administration, and general trust in a political party (see Appendix B for the exact wording). We have also three questions measuring cognitive reasoning style, and we measure charitable giving using an incentivized donation task.¹⁹ At the end of the questionnaire, we ask decision makers for their political orientation, their vote during the Brexit referendum, and their income level. We have also access to information about their gender and age through the Prolific platform.

¹⁸We use continuous sliders with ticks, but no labels except for their endpoints which read “Strongly Disagree” and “Strongly Agree”.

¹⁹Each participant had a budget of £1 and decided how much of that amount to donate to Macmillan Cancer Support. Each participant would then keep the residual. We chose the charity such that it was highly trusted by the Britons independent of political orientation and age.

2.2.5 Stage 5: Determination of payoff relevant decision

In the fifth stage, the decision relevant to payment to the anonymous receivers is determined. Note that, by design, the decision nodes that participants reach are a consequence of previous decisions. It is thus of crucial importance not to provide any strategic incentives to deviate from responding according to personal preferences, for example by purposefully stating a personal conflict in order to increase the chance for this pair to be randomly drawn. We implement a procedure that excludes such incentive effects. Instead of drawing among the decisions that decision makers actually make, we randomly select one of all potentially available decision tasks. If that decision was never reached by the decision maker we use their Stage 2 answer of the same distributional pair to determine the distribution to be implemented. Technically, we start by drawing the relevant distributional pair right after assigning the principle at the beginning of the experiment. For each of the eight pairs, the choice could be made based on principle subscription, anti-principle subscription, distributive choice, or the revised versions of these. This yields a total of $8 \times 3 \times 2 = 48$ potential decisions which are all equally likely to be drawn. For each pair, we are guaranteed to observe a distributive choice which serves as the payoff-relevant decision in case the randomly drawn decision is not applicable.²⁰ Admittedly, this procedure puts more weight on the Stage 2 decisions. However, any procedure that assigns uniform weight to all decisions a participant actually made in the experiment would need to adjust probabilities conditional on principle subscription and personal conflicts. Consequently, the above-mentioned concern regarding strategic deviation from individual preferences would become relevant. Instead, we have made sure that each decision in our experiment has a positive probability to be randomly selected, and that there are no strategic incentives to deviate from personal preferences. Note that we did not communicate the details of this procedure to participants. Instead, we informed them about the properties of the procedure, namely the fact that each decision has a positive probability to be chosen for implementation, and that there are no incentives to deviate from choosing their preferred option in any decision.

2.3 Experimental implementation details

The experiment was programmed in oTree (Chen et al., 2016) and implemented on the platform Prolific. Given our pre-registered power calculations, we invited a (gender-balanced) sample of 2,295 participants. We restricted participation to subjects who reside in the UK and are fluent in English and aimed to balance the number of observations across treatments.

²⁰The randomly drawn decision could be inapplicable because participants did not subscribe to a principle or anti-principle but opted for having their manual decision implemented. Furthermore, the random draw could select a decision from a revision stage that never occurred.

Most decision makers spent about ten to fifteen minutes on our experiment and received between £2 or £3 for participation depending on their charity choice. The average earnings were 2.63 GBP. We ran our sessions on August 27th, October 2nd, and October 3rd, 2022. Participants could read all instructions (in Appendix B) at their own pace through all stages.

3 Results

Our analysis starts with describing subscription rates, distributional choices, and how decision-makers choose to resolve any personal conflict when confronted with those (Socratic method). We then turn to study how the Socratic method causally affects the polarization of views on societal policies and distributive principles. We have pre-registered most of the research hypotheses that we address²¹ and indicate clearly when conducting exploratory analysis beyond our pre-registered analysis plan. Pre-registered analysis not presented in the main body of the paper can be found in Appendix A for completeness.

3.1 Subscriptions, distributive choices, and the resolution of personal conflict

Figure 2 shows the Stage 1 subscription rates of participants for both principles and anti-principles, pooling both treatments.²² We note that subscription rates for principles lie in the interval between 0.4 and 0.6 for all but the Gini principle. The Gini is significantly less popular than Rawls, Utilitarian, Efficient transfer (two-sided t -test, $p < 0.001$; two-sided MWU-test, $p < 0.001$), and Leaky-bucket transfer (two-sided t -test, $p = 0.004$; two-sided MWU-test, $p = 0.008$). Leaky-bucket transfer is less popular than Efficient transfer (two-sided t -test, $p = 0.021$; two-sided MWU-test, $p = 0.042$). We do not observe a statistically significant difference in Stage 1 subscriptions in any other pairwise comparison.

Notice also that each of the anti-principles (apart from anti-Gini) is significantly less subscribed to than the respective principle, with subscription rates between 0.05 and 0.15. As the principles and anti-principles were on an equal footing in our design, this strongly suggests that the subscriptions to the principles are not driven by experimenter demand.

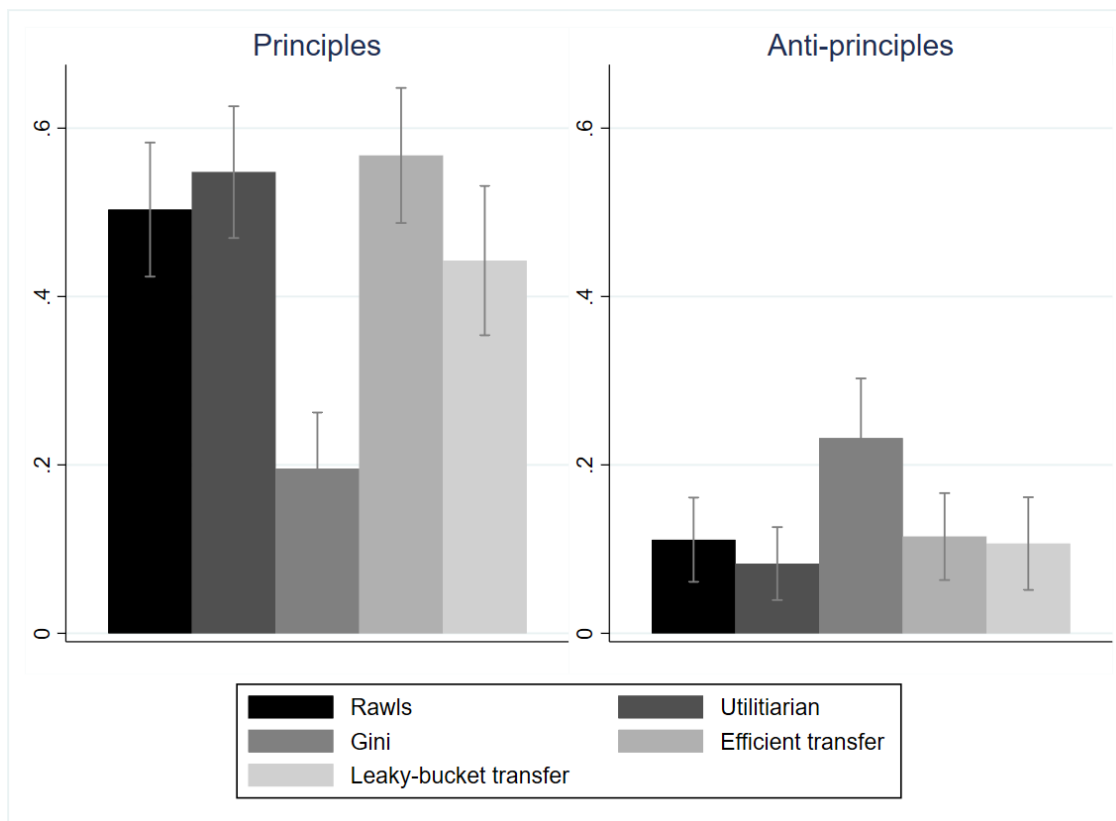
The anti-Gini principle receives a significantly higher share of subscriptions (two-sided t -test, $p < 0.016$; two-sided MWU-test, $p < 0.031$) than other anti-principles, however. In fact, Stage 1 subscription rate to anti-Gini is not significantly lower than that of the corresponding

²¹See *The Open Science Framework*, osf.io/ycd9f.

²²There are no statistically significant differences in Stage 1 subscription rates between Baseline and Socratic treatment (two-sided t -tests, $p > 0.1$; two-sided MWU-tests, $p > 0.1$), indicating that randomization to treatment worked.

Gini principle. We suspect that this is due to the complexity of the description of the Gini principle, which is generally not considered a social choice principle.²³ There is no statistically significant difference between any of the other anti-principles.

Figure 2: Stage 1 subscriptions to principles and anti-principles



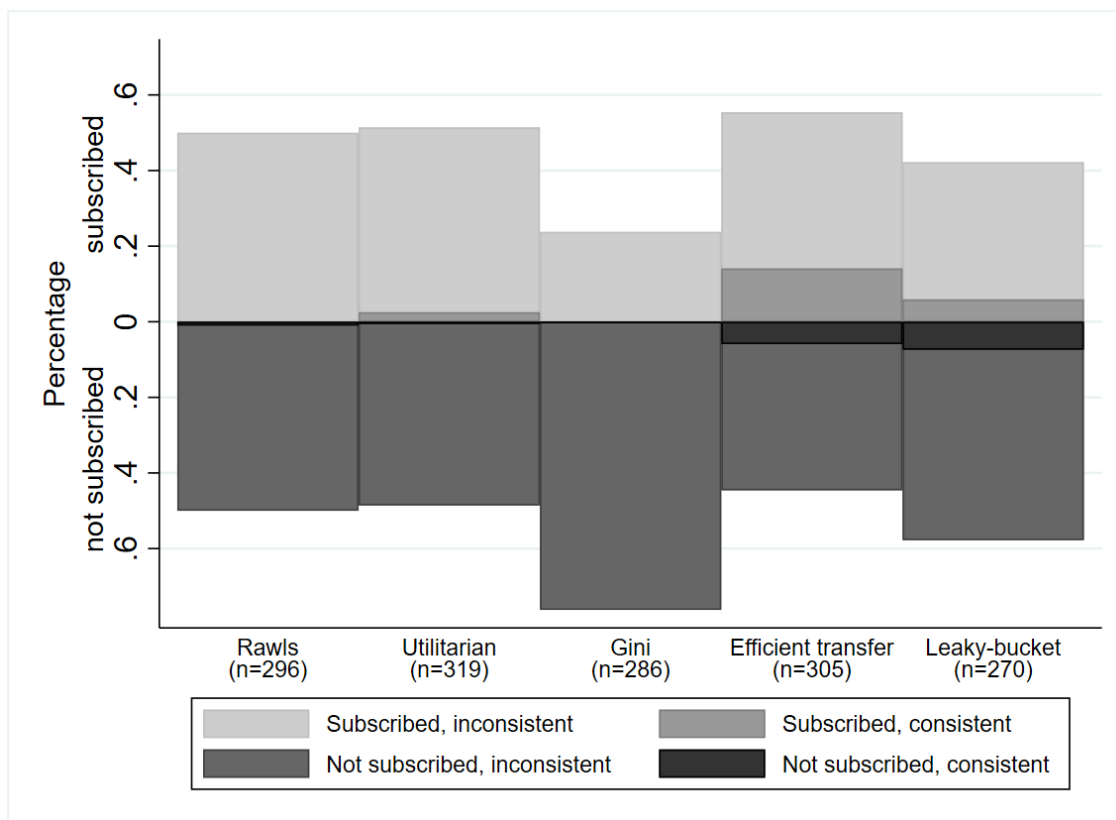
Note: We show pooled subscription rates for principles (left) and anti-principles (right), i.e., include observations from both Baseline and Socratic treatment.

In Figure 3 we have added information on distributional choices from Stage 2, across both treatments. Note that each bar has a total height of exactly one. The part below the horizontal zero line captures the share of participants who did not subscribe to the particular principle, whereas the part above the zero line captures the share who did subscribe to the principle in Stage 1. Each bar's light grey area depicts the share of participants whose distributive choices violate subscriptions to the respective principle, i.e., who reveal a personal conflict. The slightly darker grey area above the zero line depicts the share of participants who choose consistently with the principle. The darkest grey area right below the zero line depicts the share of participants who choose consistently with the principle the participant was assigned to, even if the participant did not subscribe to the principle.

²³In fact, Gini was introduced to our design due to its applied popularity in research when evaluating the equality of the distribution of income or wealth.

As is evident, Efficient transfer has the highest rate of participants with consistent choices, both with and without a subscription, about 20% in total. For other principles, inconsistent choices are abundant.

Figure 3: Stage 1 subscriptions and Stage 2 distributive choices

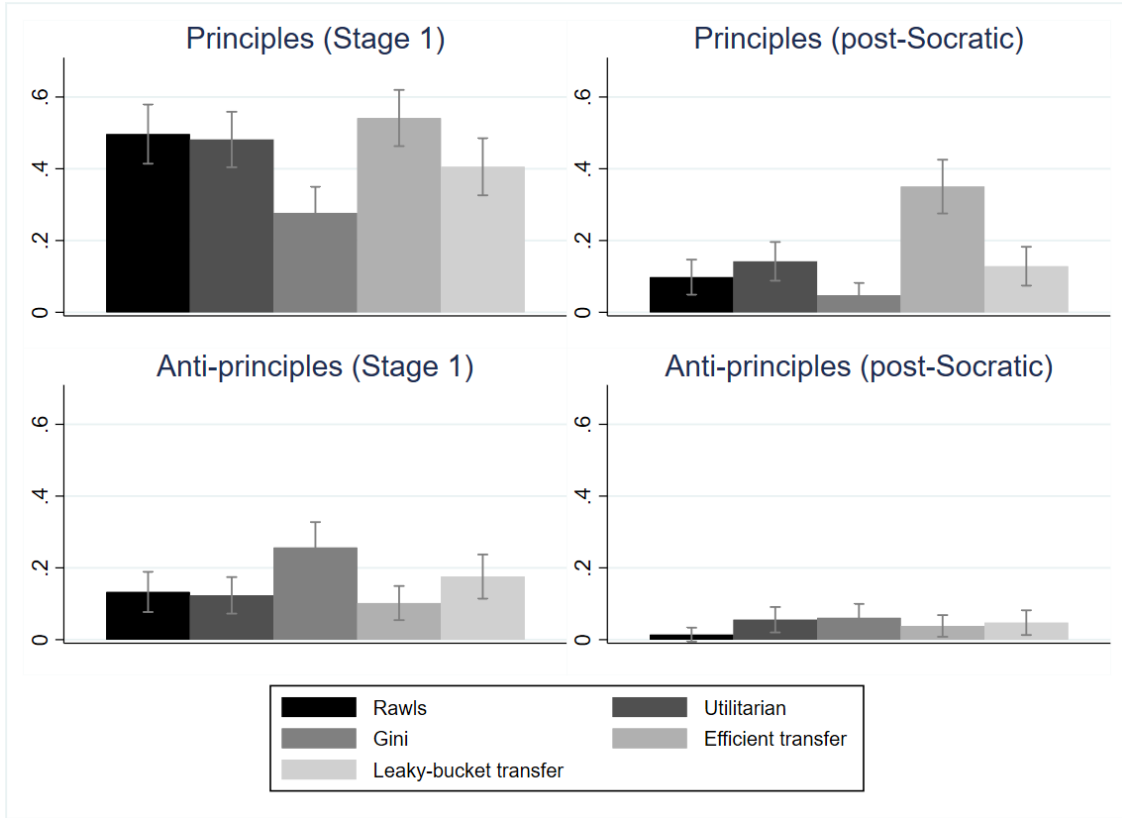


Note: Each bar has a total height of exactly one, but is shifted below zero depending on the share of participants who did not subscribe to the particular principle. Each bar is partitioned into segments which represent the share of individuals who took specific actions during Stages 1 and 2. In particular, the segment *Subscribed, inconsistent* represents individuals who subscribed to the principle in the first stage, but chose inconsistent with the principle at least once during the second stage. *Subscribed, consistent* depicts the share of individuals who subscribed to the principle and chose consistently with it throughout the second stage. The segment *Not subscribed, inconsistent* captures participants who neither subscribed nor chose consistently with the principle during Stage 2. Finally, those who did not subscribe to the principle while choosing consistently with the principle throughout all distributive choices in Stage 2 are captured in the segment *Not subscribed, consistent*.

We now describe the effect of being exposed to inconsistencies between Stage 1 and Stage 2 choices on the subscriptions to principles, i.e. Stage 3 decisions. For that purpose, it makes sense to only focus on the participants in the Socratic treatment. The top row of Figure 4 depicts Stage 1 subscriptions post-Socratic subscriptions by principle, while the bottom row depicts the corresponding subscriptions by anti-principle.

When focusing on ex-post popularity we observe that, yet again, Gini is less popular

Figure 4: Subscription rates in the Socratic treatment



Note: This figure only includes participants in our Socratic treatment. The top row depicts subscription to principles, while the bottom row refers to anti-principles. The columns show Stage 1 subscriptions and post-Socratic subscriptions, respectively. Post-Socratic subscription is defined as subscribing to the principle in Stage 1 and never changing subscription during Stage 3, or not entering Stage 3 because choices are consistent with the principle throughout Stage 2.

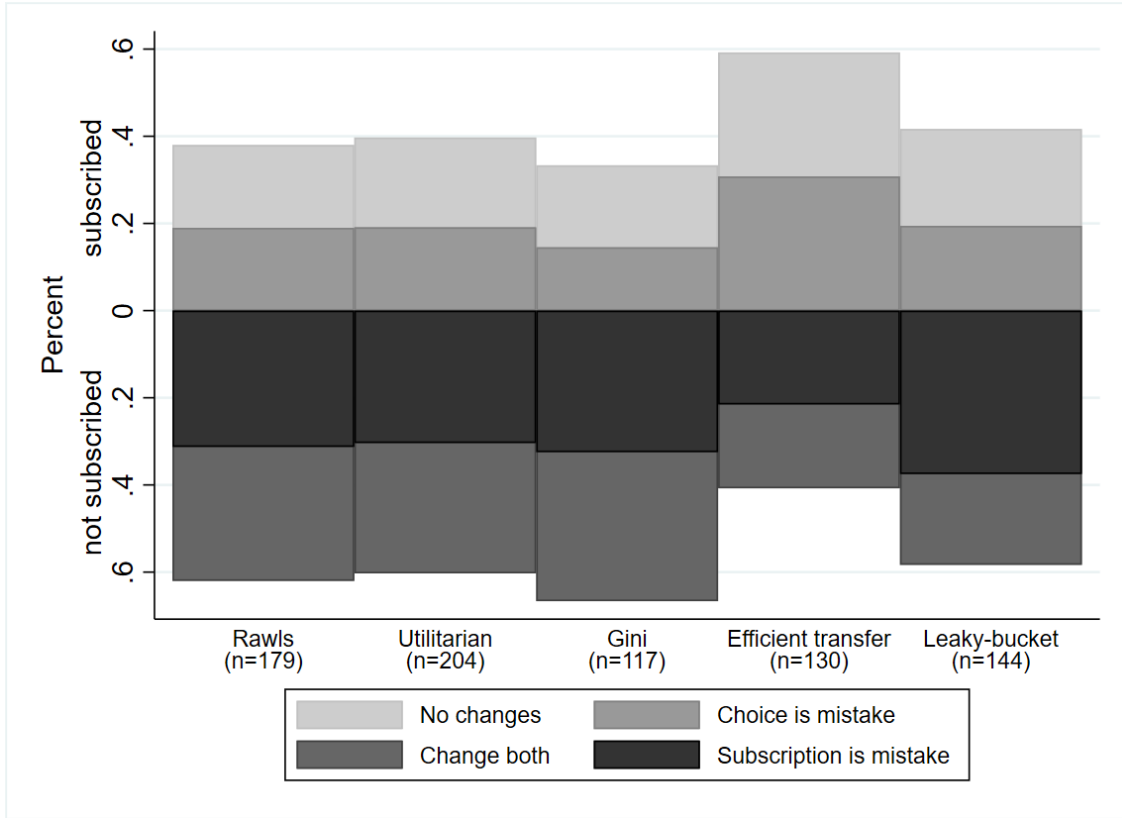
than the other principles (two-sided t -test, $p < 0.016$; two-sided MWU-test, $p < 0.032$). The central finding of this analysis is that the Efficient transfer is clearly more popular than any other principle after Stage 3 (two-sided t -test, $p < 0.001$; two-sided MWU-test, $p < 0.001$). In this sense, it is the only principle that is robust to the resolution of personal conflicts between the subscription to the principle and the distributive choices, i.e. what we call the Socratic treatment.

All principles have significantly fewer subscribers after the revision stage (one-sided t -test, $p < 0.001$; one-sided MWU-test, $p < 0.001$) compared to before. The same holds for all anti-principles (one-sided t -test, $p < 0.017$; one-sided MWU-test, $p < 0.026$).²⁴

Figure 5 describes the decisions at the revision stage, principle by principle. Each bar

²⁴After Stage 3, anti-Rawlsian has a lower subscription rate than anti-Utilitarian and anti-Gini (two-sided t -test, $p < 0.025$; two-sided MWU-test, $p < 0.050$). There are no differences in any other pairwise comparison.

Figure 5: Resolution of individual conflict in the Socratic treatment



Note: Each bar has a total height of exactly one, but is shifted below zero depending on the share of unsubscriptions from the particular principle during resolution decisions. Each bar is partitioned into segments that represent the share of specific actions during Stage 3. Note that all individual conflicts require that individuals had subscribed to the principle in Stage 1 and chose inconsistently in Stage 2. The segment *No mistake* represents resolution decisions where the decision maker neither changes the subscription to the principle, nor the distributional choice which is inconsistent with the principle. *Choice is mistake* depicts the share of decisions in which individuals did not change their subscription during the resolution stage, but aligned their distributional choice. In other words, these individuals submitted consistent decisions during the resolution stage in favour of the principle. If individuals unsubscribe during the resolution stage and keep their distributional choice which is inconsistent with the principle, they appear in segment *Subscription is mistake*. Finally, the segment *Unsubscribe and change choice* captures the remaining participants, i.e. those who change their distributional choice to be consistent with the principle, while at the same time unsubscribing from it.

above (below) the zero line depicts the share of Stage 3 *decisions* in which decision makers keep on (gave up) subscribing to the principle.²⁵ The lighter grey part right above zero line captures the choices of reversing the distributive choice to be consistent with the principle. Thus, as in Nielsen and Rehbeck (2022), these can be interpreted as cases where distributive *choices are mistakes*. The revisions in the darkest grey area right below the zero line depict the share of Stage 3 decisions in which participants unsubscribe from the respective principles

²⁵Remark the difference to Figures 2 to 4 where the bars represent the share of participants.

while sticking to the distributive choice made. These can be interpreted as cases where *subscriptions are mistakes*. When comparing these two categories we find that, except for Efficient transfer, principle subscription is revised more frequently than the distributional choice. This reaction to personal conflict contrasts with the findings of Nielsen and Rehbeck (2022), who observe that a vast majority revise their choices and conclude that generally choices are mistakes. After resolution of personal conflicts, they still find subscription rates of 85% and above for their 12 individual rational choice principles, and very few participants who would unsubscribe from a principle when facing a personal conflict. Indeed, as we stated before, the only exceptional social choice principle is Efficient transfer. The distributive choices of Efficient transfer subscribers are remarkably consistent with the subscription, significantly more so than for any other principle. An equally high fraction have conflicts but do not unsubscribe from the Efficient transfer principle. This is in stark contrast with other principles.

3.2 Causal effects of Socratic method

Let us then turn attention to the causal effects of applying the Socratic method, i.e. being informed of a personal conflict, and being allowed to resolve it. At the societal level, we look at the effects on societal conflict of views (polarization) regarding (i) subscriptions to the principles, and (ii) policy preferences.²⁶ We then turn to the individual level. In the previous section, we already learned a great deal about how people revise their personal conflicts, and how that depends on the principle randomly assigned to. In addition, we will learn how *policy preferences* change at the individual level in causal response to the Socratic treatment.

The causal effect of the resolution for the support of each principle was analyzed in the previous subsection. Since Stage 1 subscription rates are close to 50% (where the societal variation in subscriptions is maximal) and the subscription rates are drastically reduced by 15 %-points or more, there is much less societal variation in the post-Socratic subscription rates. Thus, we find an effect on the dispersion of support for each principle. The causal effects of resolution on the support of the principles is summarized in Table 1.

Table 2 summarizes societal conflict regarding policy preferences. Notice that these are the overall results summarizing effects over various principles. As one can note, indeed the Socratic method causally reduces the variation in opinions for each of the policies. The reduction is statistically significant, however, for redistributive policy only.²⁷

²⁶See section 2.2.4 for the list of policy preferences included in our study.

²⁷Note that we test for differences using Levene’s test for equal variance across treatments. Our preregistered plan was to compare pairwise differences across treatments but it turned out that this test was not suitable for our purposes.

Table 1: Societal disagreement on principles before and after treatment

	Mean		Standard deviation		Levene's p-value	N
	Stage 1	post-Socratic	Stage 1	post-Socratic		
Rawls	0.497	0.098	0.502	0.298	0.000	143
Utilitarian	0.481	0.142	0.501	0.350	0.000	162
Gini	0.277	0.047	0.449	0.213	0.000	148
Efficient transfer	0.541	0.350	0.500	0.479	0.001	157
Leaky-bucket transfer	0.405	0.128	0.493	0.336	0.000	148

Note: For each principle, we report the mean subscription rates from Stage 1 and after Stage 3, as well as the corresponding standard deviations. Finally, we report the results of Levene's test of equality of variance when comparing subscription rates from Stage 1 and after Stage 3. All p-values are rounded to 3 digit precision.

Table 2: Societal disagreement on policy preferences by treatment

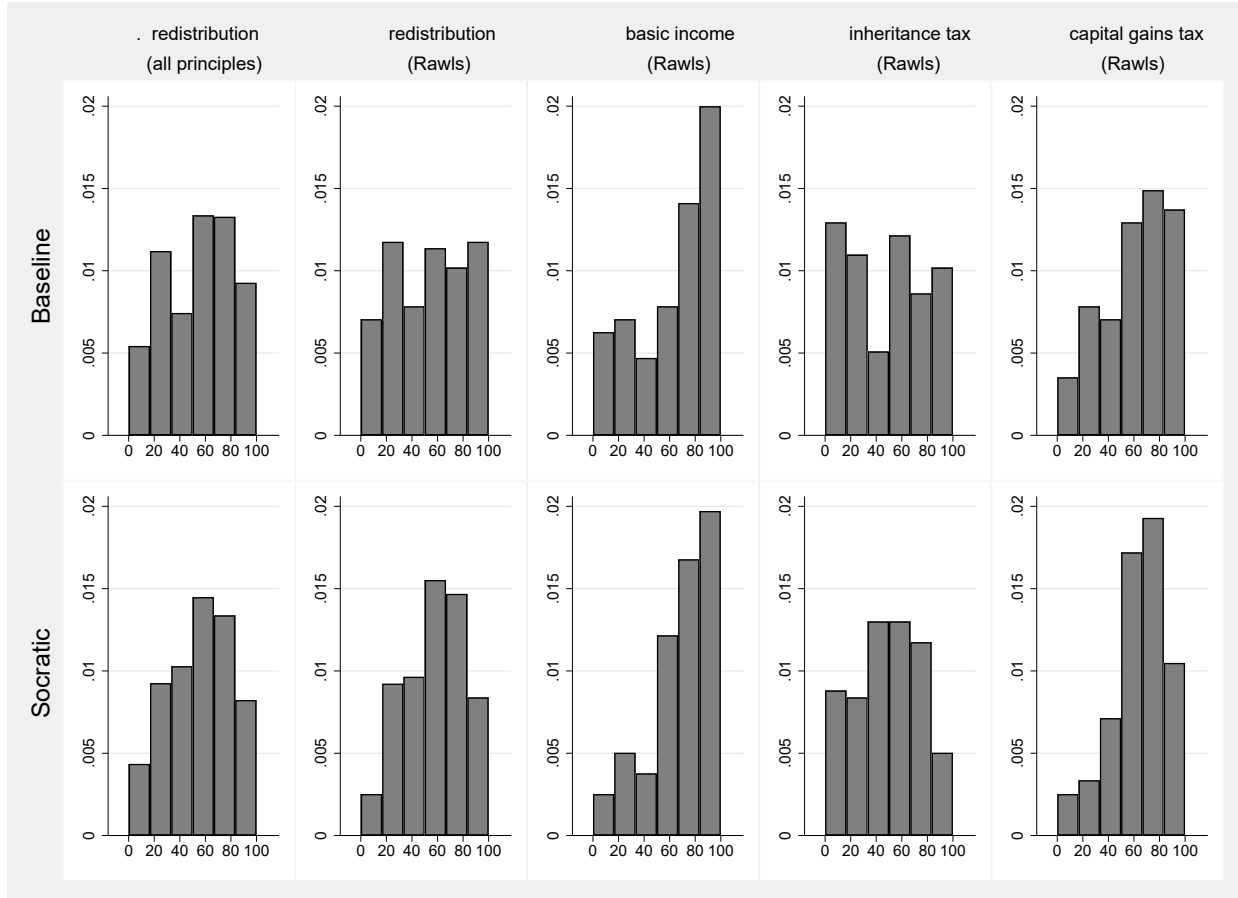
	Mean		Standard deviation		Levene's p-value	N
	Baseline	Socratic	Baseline	Socratic		
basic income	63.978	65.292	30.499	29.451	0.179	1476
early-life education	48.942	48.983	23.721	22.932	0.361	1476
redistribution	54.340	55.388	27.571	25.947	0.023	1476
progressive income tax	47.852	47.044	25.201	24.170	0.113	1476
inheritance tax	47.820	48.830	29.440	28.748	0.344	1476
capital gains tax	60.182	62.355	24.499	23.007	0.063	1476

Note: For each policy (listed in the first column), the second and the third column report the mean support in the Baseline and Socratic treatment, respectively. The dispersion of support, i.e. the societal conflict or polarization, is measured by the standard deviation in columns four and five for Baseline and Socratic treatment, respectively. In the last column, we report the results of Levene's test of equality of variance when comparing Baseline and Socratic treatment for each policy.

To better understand the effect of resolution on the support for redistributive policies, let us look at Figure 6. Each decision maker rated their support for each policy on a scale from 0 to 100. The top-left panel summarizes the distribution of support for redistributive policy in the Baseline treatment. The bottom-left panel summarizes the distribution of support in the Socratic treatment. As is evident, the resolution treatment decreases extreme opinions and shifts opinions towards the mean of the distribution. Thus, the Socratic method reduces societal conflict in opinions about redistributive policy.

We saw in the previous subsection that, for principles other than Efficient transfer, subscriptions are considered as mistakes. The individual support for the principles is reduced as a consequence of the Socratic method. How is then the individual support for the societal

Figure 6: Policy preferences by treatment



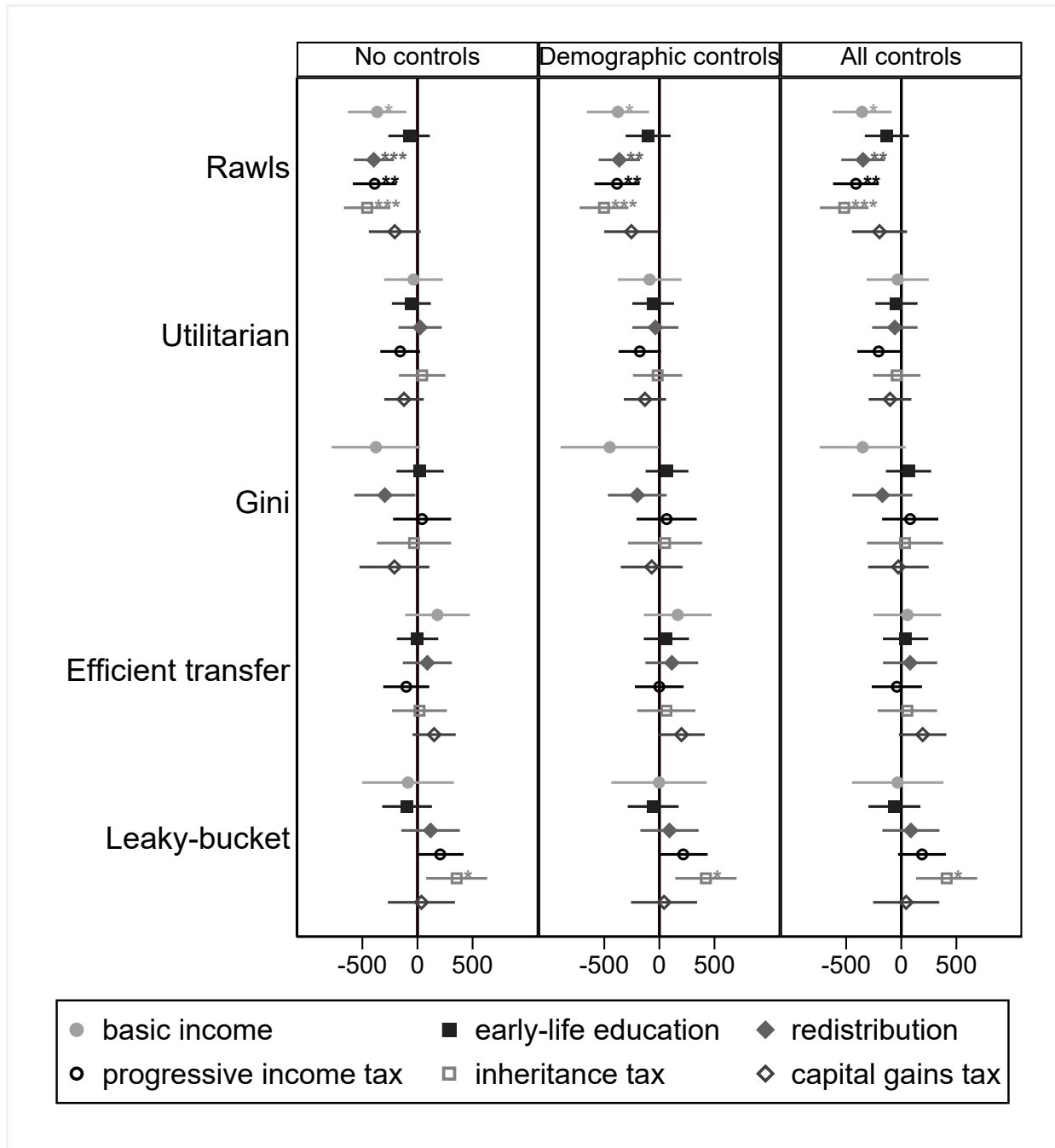
Note: We present histograms of policy preferences our decision makers reported during the survey stage. Preferences range from 0 (strongly disagree) to 100 (strongly agree). The first row shows the distributions for all decision makers in the Baseline treatment, irrespective of whether they had revealed personal conflict or not. The second row shows the distributions for all decision makers in the Socratic treatment, regardless of whether they went through the resolution stage or not. The first column depicts the distribution of preference on redistribution independent from which principle decision makers had been assigned. Columns two to five show the distributions of preferences on policies exclusively for decision makers who had been assigned the Rawlsian principle.

policies affected by the Socratic method? In Figure 7, we summarize these effects by principle and by policy. The dependent variable in the underlying regressions is the difference in individual opinion from the average opinion in society.²⁸ Consistent with the two leftmost panels in Figure A.7, we find that distance of own opinion to the average opinion is reduced by the Socratic method. This is particularly significantly visible when the Socratic method is applied to personal conflicts concerning the Rawlsian principle.

Indeed, even at the societal level, the causal effects of the Socratic method on societal conflict are particularly strong and significant for those randomly assigned to the Rawlsian

²⁸Note that this regression was not part of our pre-analysis plan.

Figure 7: Effects of resolution on quadratic deviations from mean policy preferences by principle



Note: The graph depicts marginal effects of resolution on individual quadratic deviation from mean policy preferences. Effects are estimated using linear probability models including different sets of controls. Demographic controls are age, gender, income level and CRT score. In the right column we include our full set of control variables.

Table 3: Societal disagreement on policy preferences by treatment (Rawls)

	Mean		Standard deviation		Levene's p-value	N
	Baseline	Socratic	Baseline	Socratic		
basic income	64.444	69.238	31.721	26.277	0.002	296
early-life education	51.346	47.140	25.382	23.720	0.210	296
redistribution	52.725	57.881	29.096	25.125	0.011	296
progressive income tax	48.699	46.490	26.716	23.848	0.081	296
inheritance tax	47.471	48.315	31.129	27.707	0.035	296
capital gains tax	61.222	63.392	26.700	23.887	0.049	296

principle. Table 3, and the remaining panels of Figure 6 summarize these effects. The societal conflict in opinions regarding redistributive policy, basic income, inheritance tax, and capital gains tax is reduced significantly when personal conflicts with the Rawlsian principle are revealed to the decision maker and she/he is allowed to revise her/his subscription to the principle.

4 Discussion

Our design is inspired by the research of Nielsen and Rehbeck (2022) on rationality principles in individual decision-making. They find that rationality axioms are widely subscribed to and they are robust to the resolution of personal conflicts. In their setting, choices are considered as mistakes and prescriptive theory fulfills the consultative role defined by Gilboa and Schmeidler (2001) and Gilboa (2010).

Our findings in the domain of distributive choices contrast sharply with the findings of Nielsen and Rehbeck (2022). It is yet perhaps not that surprising that agreement on rationality axioms is generally higher than agreement on social choice principles. Firstly, rationality axioms govern individual decision-making and guard the decision maker against money pumps, for instance. Secondly, they are typically not mutually conflicting but regard different aspects of the decision-making. Therefore, we argue, they are more likely to be universal and uncontroversial across different contexts and cultures. In contrast, social choice principles govern collective decision-making and are subject to more debate and controversy. These principles often involve trade-offs in the allocation of resources, liberties, or opportunities between the interests and values of various individuals. For example, the application of Rawls' maximin principle may require sacrificing some individual liberties for the sake of greater social and economic equality or insurance. This may not be acceptable to everyone, especially those whose liberties are being compromised or who put more value on incentives

to seize opportunities to generate economic surplus.²⁹ Similarly, the application of Utilitarianism may require sacrificing the interests of some individuals for the sake of greater overall efficiency, which may also be controversial. Moreover, tacitly or explicitly, social choice principles govern discussions regarding collective decision-making. Their acceptance depends on the specific values and preferences of the individuals and the dynamics of the debate. Different societies and groups may have different values, and therefore may have different social choice principles that they consider important or relevant. For example, Rawls’s maximin principle may be more relevant to some groups whereas the efficiency principle (an extreme form of Utilitarianism) may suit others, depending on their cultural, social, and economic contexts (Amiel et al., 2009). In sum, we think that the arguments put forward here explain the much lower subscription rate to social choice principles we observe.

Relating to the recent literature striving to understand how redistributive-policy preferences are causally affected by information provision (Cruces et al., 2013; Kuziemko et al., 2015; Karadja et al., 2017; Fehr et al., 2022; Bastani and Waldenström, 2021; Dunaiski and Tukiainen, 2023), we find that the Socratic method causally reduces polarization in both distributive and societal policy preferences. In particular, we find that the policy preferences are significantly less polarized, and thus current elicitation methods may underestimate the common support for economic and societal policies. The finding that the Socratic method reduces societal conflict in opinions may sound surprising. Yet, it is consistent with the findings of experimental group deliberation research in economics and political science. Ferreira et al. (2023), for instance, find that group deliberation in randomly assigned groups reduces polarization of opinions relative to no deliberation control. Grönlund et al. (2021) find that mediated deliberation in like-minded groups shifts the opinions towards the population average whereas the opposite is true without mediation. One should yet note that the mechanisms which reduce polarization in group deliberation are likely to be different from those operating in the resolution of personal conflict. Ferreira et al. (2023) find substantive evidence that the willingness to conform to the views of the group one identifies with is a likely driver of the effect of group deliberation. That channel is mute, however, in the resolution of personal conflict studied here. Rather, the decision maker might seek a reflective equilibrium (Amiel and Cowell, 1999; Gaertner and Schokkaert, 2012; Schönegger and Grodeck, 2022) where dissonance between one’s choices is reduced (Akerlof and Dickens, 1982; Konow, 2000). Overall, the findings of this paper, along with the literature on group deliberation, provide potential avenues to explore the robustness of individual preferences towards redistributive policies, and their robust elicitation.

²⁹Indeed, a caveat in the present analysis is the lack of explicit consideration of dynamic and stochastic effects of static redistribution such as the effects on incentives to innovate (Akcigit et al., 2022). Future studies could introduce both time and uncertainty more explicitly into the decision tasks.

Schokkaert and Tarrow (2021) review the empirical literature on distributive justice covering incentivized experimental studies in positive (as opposed to normative) social choice (Gaertner and Schokkaert, 2012), and representative questionnaire studies on redistributive preferences or attitudes (Alesina and Giuliano, 2011). Our findings and those of group deliberation research may have implications for this literature. Not only does group deliberation influence support for policy but, as has been shown here, even the resolution of personal conflicts at individual level reduces societal disagreement over policies. Which method then reveals the true underlying preferences over policies, let alone income distributions? Could deliberation or the resolution of personal conflict help to reduce polarization and to promote a common societal vision of societal policy? Just as correcting inconsistencies in beliefs by providing information can change the support for redistribution (Cruces et al., 2013; Karadja et al., 2017; Bastani and Waldenström, 2021), correcting inconsistencies between own principles and own choices can change support for principles and policy. Our results show, first, that although Efficient transfer may not receive majoritarian support, it is still the justice principle which is most robust to resolution of personal conflict. Moreover, we show that resolution of personal conflict can causally reduce differences in views regarding support for redistributive policies.

Our findings also relate to the burgeoning discussion on the normative analysis in behavioral economics (Camerer et al., 2003; Thaler and Sunstein, 2003; Bernheim, 2009; Fleurbaey and Schokkaert, 2013; Ambuehl et al., 2021; Bartling et al., 2023). The preponderance of revisions of choices rather than subscriptions to axioms in Nielsen and Rehbeck (2022) could be seen to support the idea that inconsistencies in choice patterns are evidence of mistakes by the individual and thus individual choice may fail to advance the (rational) self-interest. Using the terminology of Bernheim (2009, 2016), this particular behavioral revealed preference (BRP) approach claims that true preferences are revealed in the revised choices and it derives its justification from the claim that the deliberative Socratic method serves as a context where the decision maker is called upon to make a particularly informed and balanced judgment. Thus, the revised choices could be used as a basis for libertarian standard for societal policy, respecting the revealed individual judgment of well-being. Relying on non-choice data, Bernheim and Rangel (2009); Bernheim (2009, 2016) extend the BRP framework to allow deference to individual judgment even in cases where inconsistent choice data does not allow judging which choices represent true preferences, if any. Thus, Socratic choice data could be seen as a substitute to non-choice data when judging individual welfare implications. In future studies, Socratic and non-choice data could also play a complementary role.

Fleurbaey and Schokkaert (2013) extends the Bernheim and Rangel (2009) approach to the domain of distributive choices and social choice. Our study reveals a notable deviation in

the nature of distributive decisions compared to individual ones, with the former exhibiting greater levels of inconsistency. Moreover, we observe that fewer individuals modify their distributive decisions, while adherence to particular principles is more likely to be abandoned. The Efficient transfer principle is an exception: choices are more consistent with this principle and, if inconsistent, decision makers revise choices rather than subscriptions to the principle.³⁰

A form of Efficient transfer principle plays a central role in the approach of Fleurbaey and Schokkaert (2013) to produce social choice criteria which respect individual preferences and give some priority to the worse-off in the evaluation of public policy. Thus, our incomplete preference data gives some, but perhaps not unanimous support to that key assumption.

5 Conclusion

This study investigates the preference for the Rawlsian, Utilitarian-, Leaky-bucket transfer-, Efficient transfer and the minimization of the Gini coefficient distributive principles in a large sample of Britons. In addition, we study how they resolve personal conflicts (Socratic method) when preference for a principle conflicts with subsequent distributive choices over income distributions. Finally, we also study the causal effect of the Socratic method on societal disagreement about redistributive policies.

Our main results are that (i) initially, about half of our participants subscribe to distributive principles (when we exclude the minimization of Gini-coefficient which is significantly less popular than the social choice principles) and there is little variation in the popularity between the principles; (ii) nevertheless, distributive choices are not consistent with the distributive principles whether the participant subscribed to the principle or not, (iii) the majority of (incentivized) subscriptions to the justice principles are considered as mistakes by our participants when the participant is asked to resolve the personal conflict between the subscription to the distributive principle and the subsequent distributive choice. Only the prioritarian Efficient transfer transfer principle is robust to the resolution of personal conflict (Socratic method): it has a higher rate of consistent choices than any other principle whether the participant subscribed to the principle or not, the majority of revisions (Socratic method) of Efficient transfer-conflicts consider distributive choices as mistakes and keep on subscribing to the principle. The key novelty of our design is that we randomly

³⁰Since Stage 2 distributions differ depending on the principle, it is fair to ask whether such a comparison between the principles is valid, and if other distributions would produce a different result. We conjecture that one could easily find a different set of distributional pairs for each principle that yields more consistent decisions. However, we designed our decision tasks to make the likelihood of personal conflicts high. Consequently, our claim is that the inconsistencies we find may serve as an upper bound, and that Efficient transfer principle sticks out for having a lower upper bound than other principles.

assign participants to the Socratic resolution method. We find that the resolution has a causal reducing effect on the polarization of popular support for both distributive principles and societal policies.

Future work could map the effects of resolving personal conflict more broadly. For instance, one could study more complex principles, such as "Utilitarianism with a floor constraint" (Frohlich and Oppenheimer, 1992), which have received substantial empirical support in previous research (Traub et al., 2005; Gaertner and Schokkaert, 2012). Is the support robust to the resolution of personal conflict and does this latter have a causal effect on policy preferences in the case of such widely supported but more complex principles are considered?

References

- ABRAMOWITZ, A. I. AND K. L. SAUNDERS (2008): “Is polarization a myth?” *The Journal of Politics*, 70, 542–555.
- ABRAMOWITZ, A. I. AND S. WEBSTER (2016): “The rise of negative partisanship and the nationalization of US elections in the 21st century,” *Electoral Studies*, 41, 12–22.
- AKCIGIT, U., J. GRIGSBY, T. NICHOLAS, AND S. STANTCHEVA (2022): “Taxation and innovation in the twentieth century,” *The Quarterly Journal of Economics*, 137, 329–385.
- AKERLOF, G. A. AND W. T. DICKENS (1982): “The economic consequences of cognitive dissonance,” *American Economic Review*, 72, 307–319.
- ALESINA, A. AND P. GIULIANO (2011): “Preferences for redistribution,” in *Handbook of Social Economics*, Elsevier, vol. 1, 93–131.
- ALESINA, A., A. MIANO, AND S. STANTCHEVA (2023): “Immigration and redistribution,” *The Review of Economic Studies*, 90, 1–39.
- ALMÅS, I., A. W. CAPPELEN, AND B. TUNGODDEN (2020): “Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?” *Journal of Political Economy*, 128, 1753–1788.
- AMBUEHL, S., B. D. BERNHEIM, AND A. OCKENFELS (2021): “What motivates paternalism? An experimental study,” *American Economic Review*, 111, 787–830.
- AMIEL, Y. AND F. COWELL (1999): *Thinking about inequality: Personal judgment and income distributions*, Cambridge University Press.
- AMIEL, Y., F. A. COWELL, AND W. GAERTNER (2009): “To be or not to be involved: a questionnaire-experimental view on Harsanyi’s utilitarian ethics,” *Social Choice and Welfare*, 32, 299–316.
- ATKINSON, A. B. ET AL. (1970): “On the measurement of inequality,” *Journal of Economic Theory*, 2, 244–263.
- BARTLING, B., A. W. CAPPELEN, H. HERMES, AND B. TUNGODDEN (2023): “Free to fail? Paternalistic preferences in the United States,” *NHH Dept. of Economics Discussion Paper*.
- BASTANI, S. AND D. WALDENSTRÖM (2021): “Perceptions of inherited wealth and the support for inheritance taxation,” *Economica*, 88, 532–569.

- BERNHEIM, B. D. (2009): “Behavioral welfare economics,” *Journal of the European Economic Association*, 7, 267–319.
- (2016): “The good, the bad, and the ugly: A unified approach to behavioral welfare economics1,” *Journal of Benefit-Cost Analysis*, 7, 12–68.
- BERNHEIM, B. D. AND A. RANGEL (2009): “Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics,” *The Quarterly Journal of Economics*, 124, 51–104.
- BINMORE, K. G. (1998): *Game theory and the social contract: just playing*, vol. 2, MIT press.
- BINMORE, K. G. ET AL. (1994): *Game theory and the social contract: playing fair*, vol. 2, MIT press.
- BLACKORBY, C., W. BOSSERT, AND D. DONALDSON (2002): “Utilitarianism and the theory of justice,” *Handbook of Social Choice and Welfare*, 1, 543–596.
- CAMERER, C., S. ISSACHAROFF, G. LOEWENSTEIN, T. O’DONOGHUE, AND M. RABIN (2003): “Regulation for conservatives: behavioral economics and the case for” asymmetric paternalism”,” *University of Pennsylvania Law Review*, 151, 1211–1254.
- CARLSSON, F., G. GUPTA, AND O. JOHANSSON-STENMAN (2003): “Choosing from behind a veil of ignorance in India,” *Applied Economics Letters*, 10, 825–827.
- CHEN, D. L., M. SCHONGER, AND C. WICKENS (2016): “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- COHN, A., L. J. JESSEN, M. KLASNJA, AND P. SMEETS (2022): “The Wealth Gap in Fairness Preferences: Evidence from America’s Top 5%,” *Working paper*.
- CRUCES, G., R. PEREZ-TRUGLIA, AND M. TETAZ (2013): “Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment,” *Journal of Public Economics*, 98, 100–112.
- DALTON, H. (1920): “The measurement of the inequality of incomes,” *The Economic Journal*, 30, 348–361.
- DEFFAINS, B., R. ESPINOSA, AND C. THÖNI (2016): “Political self-serving bias and redistribution,” *Journal of Public Economics*, 134, 67–74.

- DETRWINKEL-KALT, M. AND M. KÖSTER (2020): “Salience and skewness preferences,” *Journal of the European Economic Association*, 18, 2057–2107.
- DUNAISKI, M. AND J. TUKIAINEN (2023): “Does income transparency affect support for redistribution? Evidence from Finland’s tax day,” Tech. rep.
- FALK, A. AND F. ZIMMERMANN (2017): “Consistency as a signal of skills,” *Management Science*, 63, 2197–2210.
- FEHR, D., J. MOLLERSTROM, AND R. PEREZ-TRUGLIA (2022): “Your place in the world: Relative income and global inequality,” *American Economic Journal: Economic Policy*, 14, 232–68.
- FEHR, E., T. EPPER, AND J. SENN (2021): “Other-regarding preferences and redistributive politics,” *University of Zurich, Department of Economics, Working Paper*.
- FERREIRA, J. V., E. SCHOKKAERT, AND B. TARROUX (2023): “How group deliberation affects individual distributional preferences: An experimental study,” Tech. rep.
- FIORINA, M. P., S. A. ABRAMS, AND J. C. POPE (2008): “Polarization in the American public: Misconceptions and misreadings,” *The Journal of Politics*, 70, 556–560.
- FISMAN, R., P. JAKIELA, AND S. KARIV (2017): “Distributional preferences and political behavior,” *Journal of Public Economics*, 155, 1–10.
- FLEURBAEY, M. AND E. SCHOKKAERT (2013): “Behavioral welfare economics and redistribution,” *American Economic Journal: Microeconomics*, 5, 180–205.
- FROHLICH, N. AND J. A. OPPENHEIMER (1992): *Choosing justice: An experimental approach to ethical theory*, vol. 22, Univ of California Press.
- FROHLICH, N., J. A. OPPENHEIMER, AND C. L. EAVEY (1987): “Choices of principles of distributive justice in experimental groups,” *American Journal of Political Science*, 606–636.
- FRYDMAN, C. AND M. MORMANN (2018): “The role of salience and attention in choice under risk: An experimental investigation,” *University of Southern California, Marshall School of Business*.
- GAERTNER, W. AND E. SCHOKKAERT (2012): *Empirical social choice: questionnaire-experimental studies on distributive justice*, Cambridge University Press.
- GILBOA, I. (2010): *Making better decisions: Decision theory in practice*, John Wiley & Sons.

- GILBOA, I. AND D. SCHMEIDLER (2001): *A theory of case-based decisions*, Cambridge University Press.
- GRÖNLUND, K. AND K. HERNE (2022): *Research Methods in Deliberative Democracy*, Oxford University Press, chap. Experimental Methods.
- GRÖNLUND, K., K. HERNE, K. STRANDBERG, AND P. SÖDERLUND (2021): “Disagreement and Deliberation: Evidence from Three Deliberative Mini-Publics,” *Political Behavior*, 1–23.
- HAMMOND, P. J. (1976): “Equity, Arrow’s conditions, and Rawls’ difference principle,” *Econometrica: Journal of the Econometric Society*, 793–804.
- HARSANYI, J. C. (1953): “Cardinal utility in welfare economics and in the theory of risk-taking,” *Journal of Political Economy*, 61, 434–435.
- (1955): “Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility,” *Journal of Political Economy*, 63, 309–321.
- (1977): “Morality and the theory of rational behavior,” *Social Research*, 623–656.
- HERNE, K. AND M. SUOJANEN (2004): “The role of information in choices over income distributions,” *Journal of Conflict Resolution*, 48, 173–193.
- HETHERINGTON, M. J. AND T. J. RUDOLPH (2015): “Why Washington won’t work,” in *Why Washington Won’t Work*, University of Chicago Press.
- HOERNER, J. M. AND S. B. HOBOLT (2020): “Unity in diversity? Polarization, issue diversity and satisfaction with democracy,” *Journal of European Public Policy*, 27, 1838–1857.
- JOHANSSON-STENMAN, O., F. CARLSSON, AND D. DARUVALA (2002): “Measuring future grandparents’ preferences for equality and relative standing,” *The Economic Journal*, 112, 362–383.
- KARADJA, M., J. MOLLERSTROM, AND D. SEIM (2017): “Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution,” *Review of Economics and Statistics*, 99, 201–212.
- KONOW, J. (2000): “Fair shares: Accountability and cognitive dissonance in allocation decisions,” *American Economic Review*, 90, 1072–1092.

- (2009): “Is fairness in the eye of the beholder? An impartial spectator analysis of justice,” *Social Choice and Welfare*, 33, 101–127.
- KUZIEMKO, I., M. I. NORTON, E. SAEZ, AND S. STANTCHEVA (2015): “How elastic are preferences for redistribution? Evidence from randomized survey experiments,” *American Economic Review*, 105, 1478–1508.
- LINDQVIST, E. AND R. ÖSTLING (2010): “Political polarization and the size of government,” *American Political Science Review*, 104, 543–565.
- NIELSEN, K. AND J. REHBECK (2022): “When choices are mistakes,” *American Economic Review*, 112, 2237–68.
- PIRTILÄ, J. AND R. UUSITALO (2010): “A ‘leaky bucket’ in the real world: estimating inequality aversion using survey data,” *Economica*, 77, 60–76.
- RAWLS, J. (1971): “A theory of justice,” *Cambridge (Mass.)*.
- ROBBETT, A. AND P. H. MATTHEWS (2018): “Partisan bias and expressive voting,” *Journal of Public Economics*, 157, 107–120.
- SCHNEIDER, M. AND J. W. LELAND (2021): “Salience and social choice,” *Experimental Economics*, 24, 1215–1241.
- SCHOKKAERT, E. AND B. TARROUX (2021): “Empirical research on ethical preferences: how popular is prioritarianism?” *GATE Working Paper Series*.
- SCHÖNEGGER, P. AND B. GRODECK (2022): “Concrete Over Abstract: Experimental Evidence of Reflective Equilibrium in Population Ethics,” In H. Viciana, F. Aguiar, & A. Gaitan (Eds.), *Issues in Experimental Moral Philosophy*.
- SCHWARDMANN, P., E. TRIPODI, AND J. J. VAN DER WEELE (2022): “Self-persuasion: Evidence from field experiments at international debating competitions,” *American Economic Review*, 112, 1118–1146.
- SEN, A. (1973): *On economic inequality*, Oxford university press.
- STARMER, C. (2000): “Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk,” *Journal of Economic Literature*, 38, 332–382.
- STRANDBERG, K., S. HIMMELROOS, AND K. GRÖNLUND (2019): “Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization,” *International Political Science Review*, 40, 41–57.

- SUGDEN, R. (2018): “The Community of Advantage: an interview with Robert Sugden,” *Erasmus Journal for Philosophy and Economics*, 13, 61–78.
- THALER, R. H. AND C. R. SUNSTEIN (2003): “Libertarian paternalism,” *American Economic Review*, 93, 175–179.
- TRAUB, S., C. SEIDL, U. SCHMIDT, AND M. V. LEVATI (2005): “Friedman, Harsanyi, Rawls, Boulding-or somebody else? An experimental investigation of distributive justice,” *Social Choice and Welfare*, 283–309.

A	Appendix	A.1
A	Supplementary Material	A.1
A.1	Additional Figures & Tables	A.2
A.2	Inconclusive preregistered analysis	A.6
B	Experimental Details & Instructions	A.9
B.1	Example screens: Subscription to principles	A.10
B.2	Example screens: Distributive choices	A.12
B.3	Example screens: Socratic method	A.15
B.4	Example screens: Policy preferences and additional controls	A.18
B.5	List of principles and survey items	A.20

A Appendix

A Supplementary Material

This section provides additional tables and figures to complement our main paper. In particular, Table A.1 in subsection A.1 shows summary statistics on each control variable of our experiment, by treatment. We do not find any statistically significant differences (except for small differences in age between our Baseline and Socratic treatment). Figure A.1 shows details on violations of principles. We provide the pairs of income distributions for each principle in Figures A.2, A.3, A.4, A.5 and A.6, respectively.

In subsection A.2 we provide the results of pre-registered analysis which we omitted in the main paper. Table A.2 and Table A.3 show the results of pre-registered regressions to investigate which individual characteristics are associated with Stage 1 subscriptions and post-Socratic subscription to principles. Given that we can not identify any interesting and robust patterns, we consider this analysis as mostly inconclusive. Similarly, we had pre-registered to study whether the process of resolution of individual conflict influence policy preferences. We show the corresponding results of overall shifts in preferences in Figure A.7. Understanding the results of our main paper, we conjecture that this analysis is likely to suffer from the fact that resolution decreases polarization by convergence towards an average opinion. If some individuals increase and others decrease their support for a given policy, an overall shift in preference is not observable.

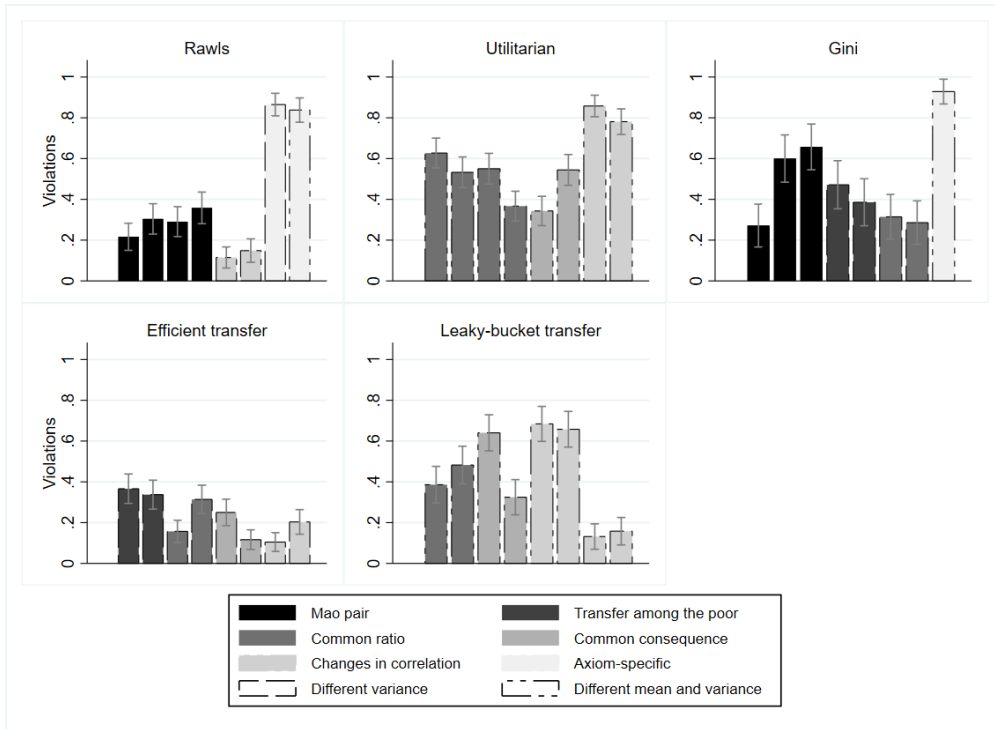
A.1 Additional Figures & Tables

Table A.1: Statistics on control variables by treatment

		Baseline			Socratic			p-value
		Mean	Std Dev	N	Mean	Std Dev	N	
Rawls	age	39.95	13.84	153	39.79	14.36	142	0.920
	female	0.44	0.50	153	0.52	0.50	143	0.170
	income level	3.43	1.29	137	3.34	1.39	128	0.567
	CRT score	1.65	1.05	153	1.73	1.05	143	0.547
Utilitarian	age	38.97	13.39	156	40.53	13.72	160	0.308
	female	0.48	0.50	157	0.53	0.50	162	0.344
	income level	3.33	1.28	139	3.47	1.26	150	0.341
	CRT score	1.56	1.06	157	1.64	1.07	162	0.529
Gini	age	41.12	14.54	138	43.07	14.89	147	0.263
	female	0.50	0.50	138	0.53	0.50	148	0.569
	income level	3.37	1.19	120	3.51	1.21	135	0.340
	CRT score	1.50	1.05	138	1.51	1.06	148	0.914
Efficient transfer	age	42.20	13.08	147	39.43	13.79	156	0.074
	female	0.53	0.50	148	0.61	0.49	157	0.170
	income level	3.45	1.20	138	3.47	1.26	142	0.879
	CRT score	1.59	1.06	148	1.50	1.12	157	0.435
Leaky-bucket transfer	age	39.34	14.34	121	41.29	13.65	148	0.255
	female	0.52	0.50	122	0.43	0.50	148	0.132
	income level	3.37	1.33	107	3.58	1.23	135	0.218
	CRT score	1.45	1.02	122	1.59	1.09	148	0.292

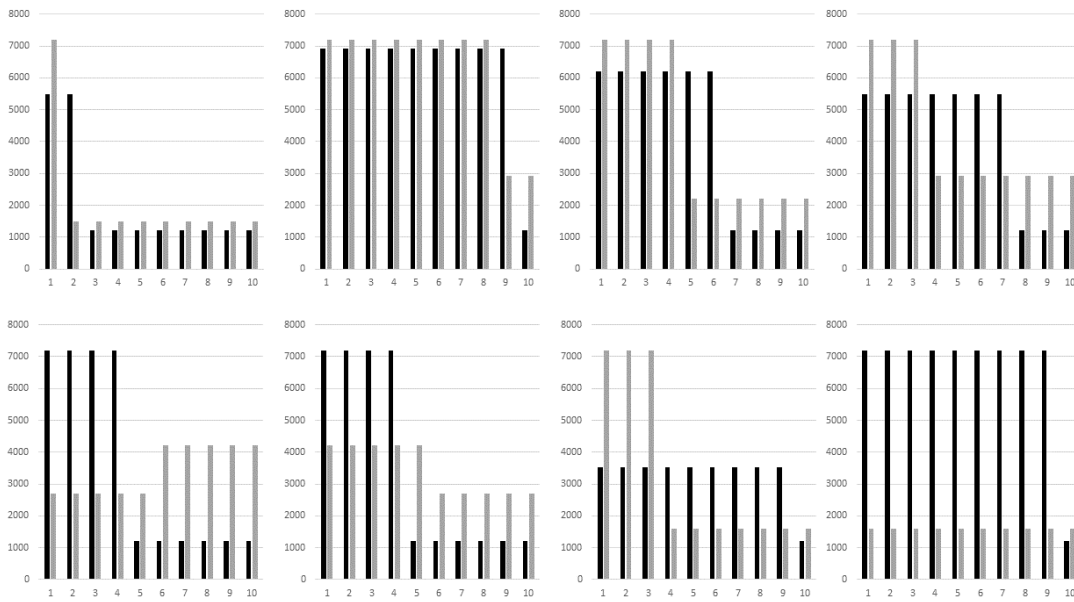
Note: We report p-values of two-sided t -tests by treatment.

Figure A.1: Share of violations by pair of income distributions



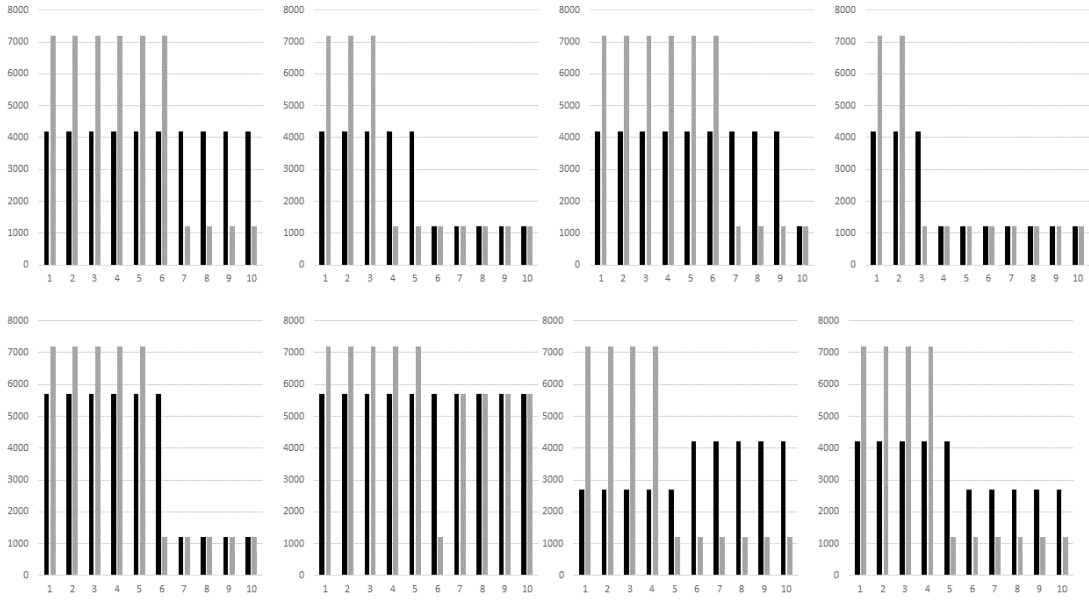
Note: We show violations by distributional pair of all decision makers who had subscribed to the respective principle.

Figure A.2: Pairs of income distributions (Rawls)



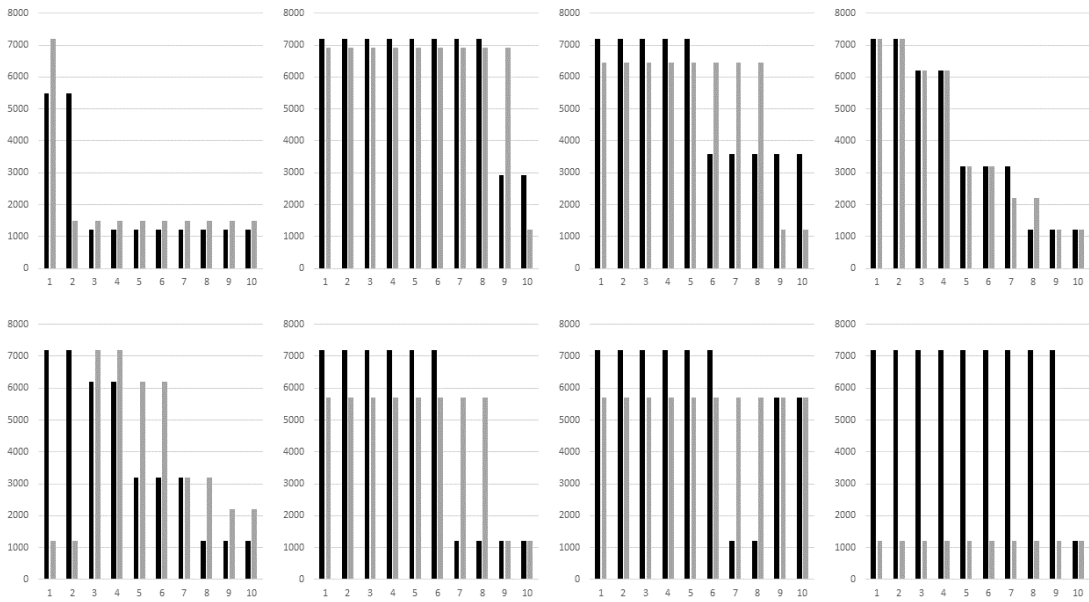
Note: We present these pairs in randomized order during stage two. We also shuffle the ten recipients randomly.

Figure A.3: Pairs of income distributions (Utilitarian)



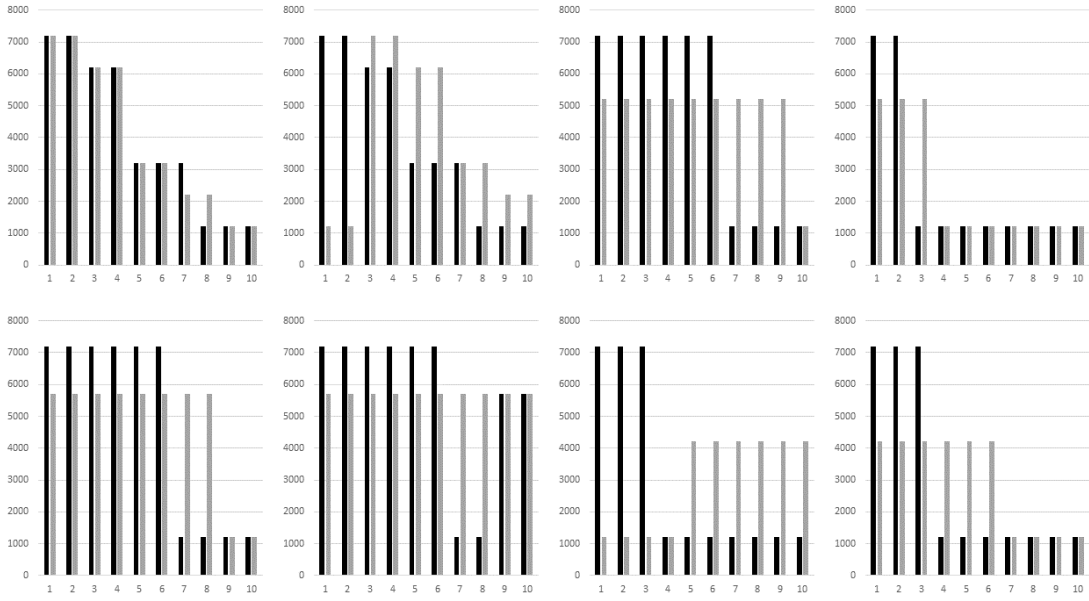
Note: We present these pairs in randomized order during stage two. We also shuffle the ten recipients randomly.

Figure A.4: Pairs of income distributions (Gini)



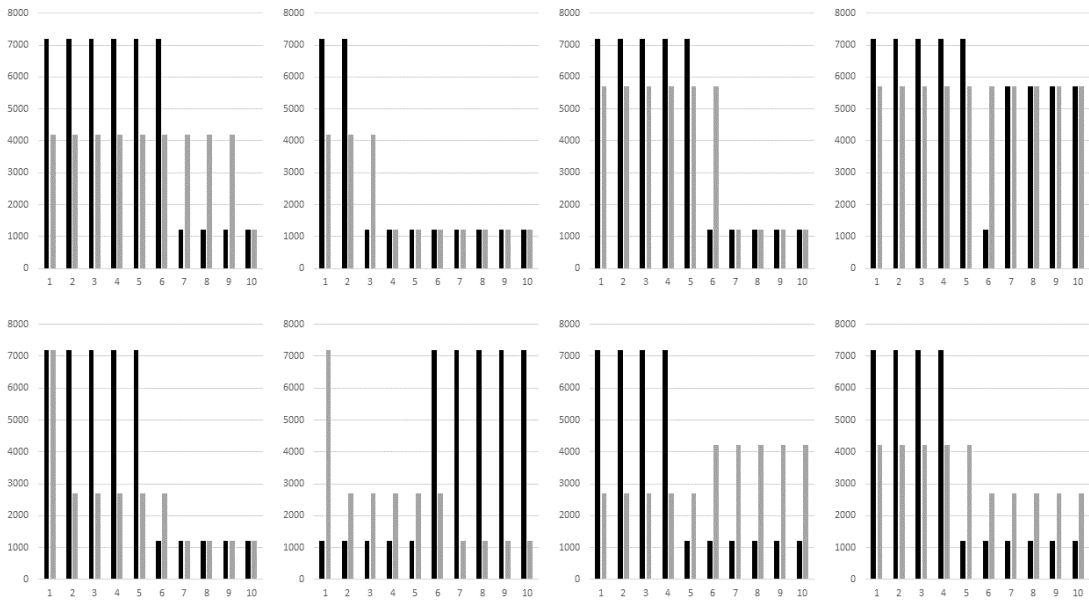
Note: We present these pairs in randomized order during stage two. We also shuffle the ten recipients randomly.

Figure A.5: Pairs of income distributions (Efficient transfer)



Note: We present these pairs in randomized order during stage two. We also shuffle the ten recipients randomly.

Figure A.6: Pairs of income distributions (Leaky-bucket transfer)



Note: We present these pairs in randomized order during stage two. We also shuffle the ten recipients randomly.

A.2 Inconclusive preregistered analysis

Table A.2: Stage 1 subscription to principles

	Rawls		Utilitarian		Gini		Efficient transfer		Leaky-bucket transfer	
	LPM	Logit	LPM	Logit	LPM	Logit	LPM	Logit	LPM	Logit
political right	0.00 (0.00)	-0.00 (0.01)	0.00 (0.00)	0.00 (0.02)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.01*** (0.00)	-0.01** (0.00)	-0.03*** (0.01)
for Brexit	0.14 (0.07)	0.63* (0.30)	0.23 (0.12)	1.03* (0.45)	-0.31 (0.04)	-2.18* (0.92)	-0.08 (0.09)	-0.38 (0.41)	0.16 (0.05)	0.87* (0.39)
against Brexit	-0.04 (0.04)	-0.19 (0.22)	0.21* (0.01)	0.92*** (0.09)	-0.08 (0.08)	-0.36 (0.48)	-0.00 (0.05)	0.01 (0.21)	-0.09 (0.22)	-0.39 (0.97)
income level	-0.01 (0.01)	-0.03 (0.03)	-0.01 (0.06)	-0.03 (0.24)	0.05 (0.01)	0.25*** (0.06)	-0.01 (0.04)	-0.06 (0.18)	-0.02 (0.01)	-0.13*** (0.01)
CRT score	0.05 (0.00)	0.22*** (0.00)	0.03 (0.04)	0.12 (0.17)	-0.02 (0.01)	-0.06*** (0.01)	0.07 (0.02)	0.34*** (0.08)	0.09*** (0.00)	0.47*** (0.12)
charitable giving	0.01 (0.06)	0.04 (0.25)	0.09 (0.13)	0.40 (0.52)	0.05 (0.10)	0.31 (0.59)	0.05 (0.10)	0.26 (0.41)	0.05 (0.01)	0.30** (0.11)
rich deserve to be rich	0.00 (0.00)	0.01 (0.02)	0.00 (0.00)	0.01** (0.00)	-0.00 (0.00)	-0.02*** (0.01)	0.00 (0.00)	0.01*** (0.00)	0.00 (0.00)	0.00* (0.00)
poor deserve to be poor	0.00 (0.00)	0.01 (0.01)	-0.00 (0.00)	-0.02*** (0.00)	0.00 (0.00)	0.02*** (0.01)	-0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.01 (0.01)
luck matters	0.00 (0.00)	0.02* (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.01*** (0.00)	0.01 (0.00)	0.03*** (0.01)	-0.00 (0.00)	-0.02*** (0.00)
taxes affect effort	-0.00 (0.00)	-0.00 (0.00)	0.00*** (0.00)	0.01*** (0.00)	-0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.01 (0.01)	-0.00 (0.00)	0.00 (0.01)
trust in party	0.00 (0.00)	0.02*** (0.00)	-0.00 (0.00)	-0.00 (0.01)	0.00 (0.00)	0.01*** (0.00)	0.00 (0.00)	0.01 (0.01)	-0.00 (0.00)	-0.02*** (0.00)
effective administration	-0.00 (0.00)	-0.02*** (0.00)	0.00 (0.00)	0.00 (0.01)	-0.00 (0.00)	-0.00 (0.01)	-0.00 (0.00)	-0.00 (0.01)	0.00 (0.00)	0.02* (0.01)
age	-0.00 (0.01)	-0.01 (0.03)	-0.00 (0.00)	-0.01* (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	-0.00 (0.00)	0.00 (0.01)
female	-0.02 (0.08)	-0.10 (0.33)	-0.10 (0.12)	-0.41 (0.50)	0.09 (0.09)	0.50 (0.45)	0.10 (0.07)	0.44 (0.31)	0.12 (0.02)	0.53*** (0.12)
constant	0.24 (0.42)	-1.26 (2.11)	0.26 (0.27)	-1.10 (1.18)	0.11 (0.15)	-1.96*** (0.53)	-0.10 (0.27)	-2.90* (1.17)	0.82 (0.13)	1.72*** (0.14)
<i>N</i>	126	126	148	148	131	131	140	140	133	133

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

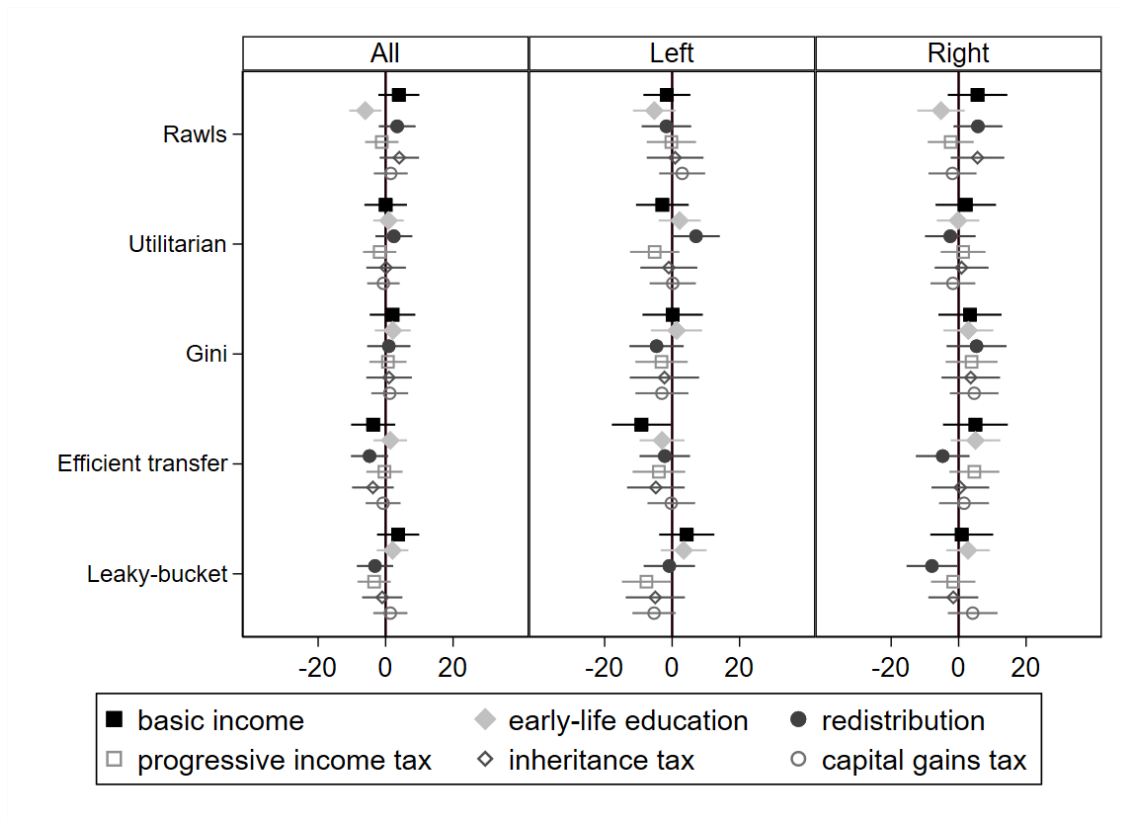
Table A.3: Post-Socratic subscription to principles

	Rawls		Utilitarian		Gini		Efficient transfer		Leaky-bucket transfer	
	LPM	Logit	LPM	Logit	LPM	Logit	LPM	Logit	LPM	Logit
political right	-0.00 (0.00)	-0.01 (0.01)	0.00 (0.00)	0.01 (0.03)	-0.00 (0.00)	-0.03*** (0.00)	0.00 (0.00)	0.00 (0.01)	-0.00 (0.00)	-0.01 (0.01)
for Brexit	-0.03 (0.17)	-0.59 (1.81)	0.09 (0.09)	0.98 (1.26)	-0.13 (0.03)	. (.)	0.03 (0.18)	0.17 (0.82)	0.15 (0.23)	0.95 (1.39)
against Brexit	-0.06 (0.08)	-0.84 (0.92)	0.16 (0.05)	1.58* (0.72)	-0.07 (0.03)	-1.78*** (0.53)	0.15 (0.10)	0.69 (0.42)	0.03 (0.02)	0.14 (0.14)
income level	0.00 (0.01)	0.05 (0.11)	-0.02 (0.02)	-0.16 (0.16)	0.01 (0.03)	0.42 (0.49)	-0.02* (0.00)	-0.12*** (0.01)	-0.02* (0.00)	-0.16*** (0.03)
CRT score	-0.01 (0.00)	-0.15 (0.11)	-0.03 (0.01)	-0.23*** (0.05)	-0.02 (0.02)	-0.28*** (0.03)	0.06 (0.03)	0.28* (0.13)	0.04* (0.00)	0.31*** (0.05)
charitable giving	-0.12 (0.02)	-1.58*** (0.13)	0.11 (0.20)	1.00 (1.24)	0.01 (0.07)	1.83 (2.49)	0.06 (0.02)	0.28*** (0.05)	-0.18 (0.11)	-1.20* (0.60)
rich deserve to be rich	0.00* (0.00)	0.02* (0.01)	0.00 (0.00)	0.02 (0.03)	-0.00* (0.00)	-0.03* (0.01)	0.00 (0.00)	0.02*** (0.00)	0.00 (0.00)	0.00 (0.01)
poor deserve to be poor	-0.00 (0.00)	-0.00 (0.01)	-0.00 (0.00)	-0.02 (0.01)	0.00 (0.00)	0.04* (0.02)	-0.00 (0.00)	-0.01 (0.01)	0.00 (0.00)	0.01 (0.01)
luck matters	0.00 (0.00)	0.00 (0.01)	-0.00 (0.00)	0.00 (0.01)	-0.00 (0.00)	-0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)
taxes affect effort	0.00 (0.00)	0.01*** (0.00)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.05 (0.04)	-0.00 (0.00)	-0.01*** (0.00)	0.00 (0.00)	0.00 (0.01)
trust in party	-0.00 (0.00)	-0.01*** (0.00)	0.00 (0.00)	0.01 (0.01)	-0.00 (0.00)	-0.07** (0.02)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.00 (0.01)
effective administration	-0.00 (0.00)	-0.00 (0.05)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.04 (0.03)	0.00 (0.00)	0.00 (0.00)	0.00* (0.00)	0.02*** (0.00)
age	0.00 (0.00)	0.04*** (0.01)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.03*** (0.00)	-0.00 (0.00)	-0.00 (0.01)	-0.00 (0.00)	-0.00 (0.02)
female	-0.06 (0.03)	-0.70 (0.77)	-0.05 (0.02)	-0.38 (0.39)	-0.01 (0.01)	0.18 (0.95)	-0.04 (0.06)	-0.20 (0.23)	0.04 (0.01)	0.25 (0.18)
constant	0.07 (0.13)	-2.75 (2.22)	-0.11* (0.01)	-4.94** (1.65)	0.11 (0.03)	-4.70 (4.65)	0.25 (0.22)	-1.10 (1.01)	0.21 (0.22)	-1.41 (1.35)
<i>N</i>	126	126	148	148	131	101	140	140	133	133

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure A.7: Effects of resolution on policy preferences by principle



Note: The graph depicts marginal effects estimated using linear probability models including controls.

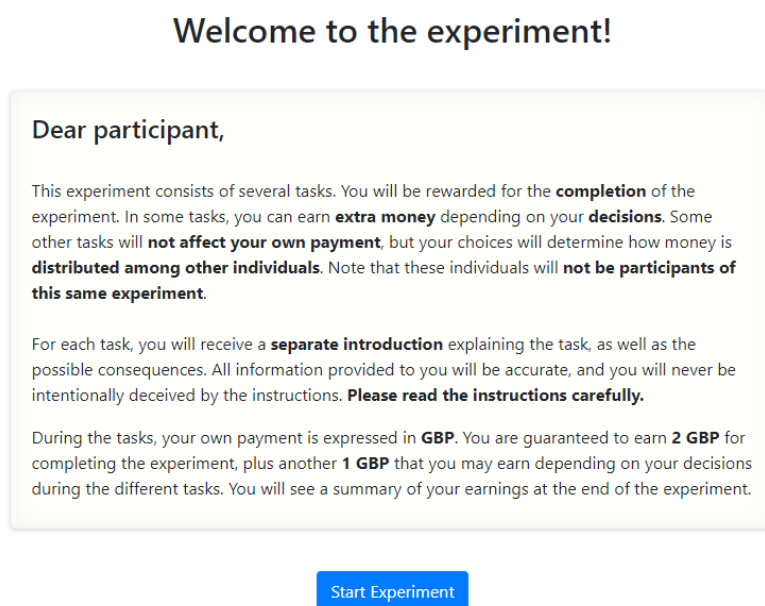
B Experimental Details & Instructions

Table A.4 summarizes the stages in our experiment. We provide screenshots from the different stages of our experiment in the following subsections. In particular, screenshots from Stage 1 are included in subsection B.1. Subsection B.2 shows example screens of what participants could see during Stage 2. In subsection B.3 we provide instructions and decision screens of Stage 3. We provide the instructions and questions of our debriefing survey at Stage 4 in subsection B.4. Finally, we provide the exact wording of each principle and survey item in subsection B.5.

Table A.4: Stages in our experiment

Stage	
0	Welcome screen
1	Subscription to principles
2	Distributive choices
3	Socratic method
4	Policy preferences and additional controls
5	Determination of payoff relevant decision

Figure A.8: Screenshot of Welcome screen



B.1 Example screens: Subscription to principles

Figure A.9: Screenshots of Stage I

Instructions

In this survey, you will be presented with different **principles** of how **income** should be **distributed** between individuals. These may be used to decide between several income programs that propose how to divide money between 10 individuals.

Your task is to **decide** which of these distributional **principles you agree with**, i.e. which principles you would like to be implemented when money is to be divided between other individuals (not involving yourself).

In particular, we will verbally describe to you two income distributions, A and B, and a principle which distinguishes between the two distributions. Subsequently, you can choose to **commit or not to the principle** (of how to distribute money to 10 randomly chosen individuals whenever it can be applied).

Here is an example:

Consider two income programs A and B such that all incomes in program A are **even numbers**, and all incomes in program B are **odd numbers**.

Principle: In any such situation, **program A** with even numbers should be chosen.

You will then have **two options** to choose from.

The first option is "**I want to commit to the principle.**" You should choose this option if you think that the principle **will always make your preferred decision** between any two distributions that **it can be applied to**, i.e. whenever all incomes of program A are even numbers, and all incomes of program B are odd numbers.

The second option is "**I do not want to commit, but decide case by case.**" You should choose this option if you think that the principle **will not always make your preferred decision** between any two distributions that **it can be applied to**, for example because you think that sometimes distribution B in which all incomes are odd numbers might be more desirable than distribution A in which all incomes are even numbers.

Your commitment decisions are **consequential**. Specifically, we will implement an income distribution for a group of 10 randomly drawn individuals based on your choices to commit or not to commit.

Figure A.10: Screenshots of Stage I continued

This is how it works:

At the end of the experiment, a **random mechanism** will determine which of your decisions is chosen for implementation. Each decision in this stage can be chosen for implementation. You should **always choose** the option that you **actually prefer**, because this is the only way to increase the chance that your preferred option is implemented.

Associated with each principle, there is a **pair of income distributions**. The two distributions differ from each other such that the **principle can be applied** in the way it was described to you. This means that it is possible to infer which distribution should be implemented based on the principle.

If you **commit** to the principle, then the **choice** between the two distributions is made **automatically based on the principle**, and this distribution will be implemented.

If you chose **not to commit**, then the pair will be presented to you to **make a decision manually**, and the distribution you choose will be implemented.

Continue

Decision

Program A is preferable compared to program B if the sum of all incomes of program A is larger than the sum of all incomes of program B.

Please choose your commitment to the principle:

- I want to commit to the principle.
- I do not want to commit, but rather decide case by case.

Submit decision

B.2 Example screens: Distributive choices

Figure A.11: Screenshots of Stage II

Instructions

In this stage of the experiment you will be asked to **choose** between **two different income distributions**, i.e. income distributions, where money is divided between **10 randomly drawn individuals**.

In particular, we will explicitly present to you a pair of income programs, A and B, using **bar charts**. For each **individual**, you will see two bars. The **first bar** (in blue) for each individual corresponds to the income that this individual would receive **under program A**. Respectively, the **second bar** (in black) corresponds to the income that the same individual would receive **under program B**.

Your task is to **choose** either **program A or program B**, depending on which of the programs generates a more preferable distribution of money over the ten other individuals (not involving yourself).

Here is an example:

Consider the following income programs A and B depicted in the bar charts below. The **first bar** for each individual corresponds to the income that this individual would receive **under program A**. Respectively, the **second bar** for each individual corresponds to the income that the same individual would receive **under program B**.

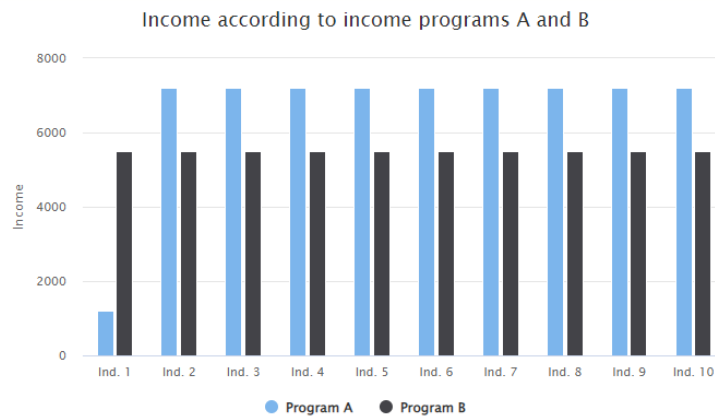


Figure A.12: Screenshots of Stage II continued

You will then have **two options** to choose from.

The first option is "**I want program A to be implemented.**" You should choose this option, if you think that distribution A is preferable out of the two distributions presented to you.

The second option is "**I want program B to be implemented.**" You should choose this option, if you think that distribution B is preferable out of the two distributions presented to you.

Your distribution choices are **consequential**. Specifically, we will implement an income distribution for a group of 10 randomly drawn individuals based on your choices.

This is how it works:

At the end of the experiment, a **random mechanism** will determine which of your decisions is chosen for implementation. Each decision can be chosen for implementation. You should **always choose** the option that you **prefer**, because this is the only way to increase the chance that your preferred option is implemented.

If you chose **program A** in the randomly drawn pair of income distributions, then program A from this pair will be implemented.

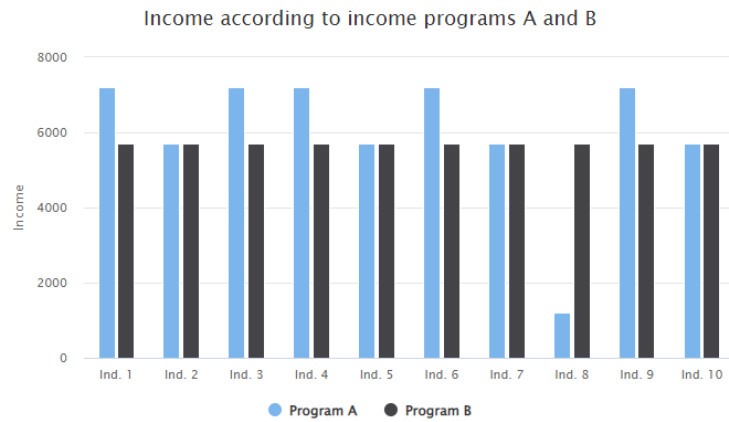
If you chose **program B** in the randomly drawn pair of income distributions, then program B from this pair will be implemented.

Continue

Figure A.13: Screenshots of Stage II continued

Choosing income program

Consider the following income programs A and B depicted in the bar charts below. The **first bar** (in blue) for each individual corresponds to the income that this individual would receive **under program A**. Respectively, the **second bar** (in black) for each individual corresponds to the income that the same individual would receive **under program B**.



Please choose which program should be implemented:

- I want program A to be implemented.
- I want program B to be implemented.

Submit

B.3 Example screens: Socratic method

Figure A.14: Screenshots of Stage III

Instructions

In this stage of the experiment you will be asked to **revise** your previous decisions in case there are **inconsistencies** between your commitment to a principle in the first stage and your decisions in the second stage.

In particular, we will explicitly present to you the **principle** that you committed to, as well as the **decision** that is **inconsistent with the commitment**. This way, you can analyze the principle and your choice together.

You will then have the opportunity to **make changes** if you wish to do so. (1) You might decide that you want to change your distribution choice to align with what the principle would have chosen for you. (2) Or you might decide that the principle should not always be applied, so you can *withdraw* from the principle and uncommit. (3) *Note that you can also stick to your decisions in the earlier stages and not make any changes, i.e. you can leave both your commitment to the principle and your distribution choice as they were.*

Again, your decisions are **consequential**. Specifically, we will implement an income distribution for a group of 10 randomly drawn individuals based on your choices.

This is how it works:

At the end of the experiment, a **random mechanism** will determine which of your decisions is chosen for implementation. Each decision can be chosen for implementation. You should **always choose** the option that you **prefer**, because this is the only way to increase the chance that your preferred option is implemented.

Again, there is a **pair of income distributions** associated with each principle (which may differ from the pair which you are given an option to revise). The two distributions differ from each other such that the **principle can be applied** in the way it was described to you. This means that it is possible to infer which distribution should be implemented based on the principle.

If the principle choice is picked by the random mechanism to be relevant and you are **committed** to the principle, then the **choice** in that pair of distributions (not necessarily the same pair which you are given an option to revise) is made **automatically based on the principle**, and this distribution will be implemented. If the principle choice is picked by the random mechanism to be relevant and you are **not committed** to the principle, then the pair will be presented to you to **make a decision manually**, and the distribution you choose will be implemented.

If the distribution choice is picked by the random mechanism to be relevant, then the payoffs are determined by the distribution choice you make. If you choose **program A** in this pair, then program B from this pair will be implemented. If you choose **program B** in this pair, then program B from this pair will be implemented.

Figure A.15: Screenshots of Stage III continued

Decision

There was an **inconsistency** regarding your commitment to a principle in the first stage and one of your decisions in the second stage. Thus, we present you both choices once more, giving you the **opportunity to make changes** if you wish to do so. Note that you can also stick to your decisions in the earlier stages and **not make any changes**.

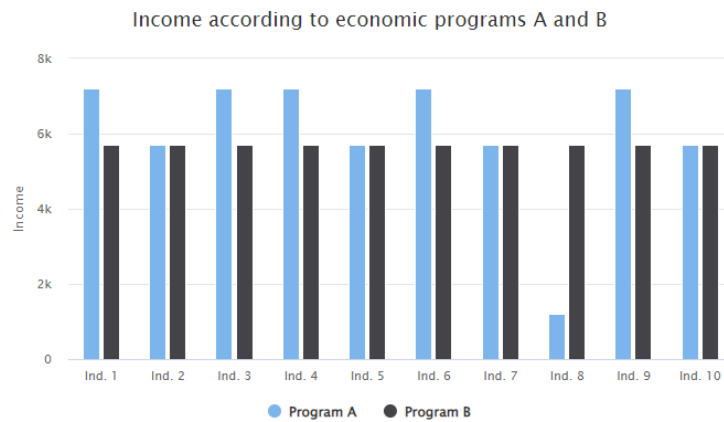
In the first stage, you were presented the following principle:

Program A is preferable compared to program B if the sum of all incomes of program A is larger than the sum of all incomes of program B.

Your choice: **I want to commit to the principle.**

In the second stage, you were presented the following decision task:

Consider the following economic programs A and B depicted in the bar charts below. The **first bar** for each individual corresponds to the income that this individual would receive **under program A**. Respectively, the **second bar** for each individual corresponds to the income that the same individual would receive **under program B**.



Your choice: **I want program B to be implemented.**

Figure A.16: Screenshots of Stage III continued

Below, you can make your final choice on whether to commit to the principle, as well as the distribution decision. Recall that **both decisions** may be randomly picked for **implementation**. Also recall that, if the principle commitment decision is randomly picked, the principle is applied to a pair of income distributions, which may differ from the pair that you are given an option to revise.

Please choose your commitment to the principle:

- I want to commit to the principle.
- I do not want to commit, but rather decide case by case.

Please choose which program should be implemented:

- I want program A to be implemented.
- I want program B to be implemented.

Submit decision

B.4 Example screens: Policy preferences and additional controls

Figure A.17: Screenshots of Stage IV

Survey

In this survey, we will ask you whether you agree or disagree with certain statements. Please use the adjustable sliders to indicate your opinion.

[Continue](#)

High income earners in our society deserve their high incomes. Do you agree with that statement?

Strongly Disagree Strongly Agree

Questionnaire

This stage is a questionnaire with three questions. Note that these questions **do not serve as attention checks**, i.e., they do **not affect your compensation** for this survey.

Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?

A man buys a pig for 60 GBP, sells it for 70 GBP, buys it back for 80 GBP, and sells it finally for 90 GBP. How much has he made?

Serena decided to invest 8,000 GBP in the stock market one day early in 2008. Six months after she invested, on July 17, the stocks she had purchased were down 50%. Fortunately for Serena, from July 17 to October 17, the stocks she had purchased went up 75%. At this point, Serena has:

[Continue](#)

Figure A.18: Screenshots of Stage IV continued

Donation

For the completion of this study you are rewarded with 2 GBP, plus a bonus of 1 GBP = 100 pence. You can now decide whether you want to donate (some of) your 100 pence bonus payment to the charity *Macmillan Cancer Support*. Any amount that you choose to donate will be sent to the charity. A receipt of total payment will be provided at a later stage. Any residual pence left after your donation will be paid to your Prolific account. Please indicate how many pence you would like to donate to charity by clicking on the slider below.

 I donate p.

Continue

Final questions (1 of 2)

You have now reached the final stage of the survey. Please remember that all your answers will be treated confidentially and no-one will be able to track them back to you as an individual. In particular, the following items will not influence your relation with Prolific or your payment in any way. Please answer the questions truthfully, as they are important for our research.

How did you vote on the EU referendum (Brexit referendum) on June 23, 2016?

It is sometimes said that political views can be placed on a left-right scale. Where would you place yourself on such a scale?

Left Right

Continue

Final questions (2 of 2)

What is your (approximate) monthly gross income (in GBP)?

Out of 100 randomly drawn citizens of the UK, how many do you think have a higher income than you?

Answer:

Continue

B.5 List of principles and survey items

Principles and their respective anti-principles:

- Rawls: Program A is preferable compared to program B if the lowest income in program A is larger than the lowest income in program B.
- anti-Rawls: Program A is preferable compared to program B if the lowest income in program A is smaller than the lowest income in program B.
- Utilitarian: Program A is preferable compared to program B if the sum of all incomes of program A is larger than the sum of all incomes of program B.
- anti-Utilitarian: Program A is preferable compared to program B if the sum of all incomes of program A is smaller than the sum of all incomes of program B.
- Gini: Consider two randomly chosen individuals in program A and two in program B. Program A is preferable compared to program B if the expected income difference between individuals (in proportion to the average income of the respective program) is smaller in program A than in B.
- anti-Gini: Consider two randomly chosen individuals in program A and two in program B. Program A is preferable compared to program B if the expected income difference between individuals (in proportion to the average income of the respective program) is larger in program A than in B.
- Efficient transfer: Program A is preferable compared to program B if there is a way to re-allocate income starting from program B and yielding program A such that re-allocations only involve transfers from individuals with higher incomes to individuals with lower incomes.
- anti-Efficient transfer: Program A is preferable compared to program B if there is a way to re-allocate income starting from program B and yielding program A such that re-allocations only involve transfers from individuals with lower incomes to individuals with higher incomes.
- Leaky-bucket transfer: Program A is preferable compared to program B if there are two or more persons such that the poorest of them earns more in A than in B, and the total earnings in A are at most as much as in B. All other persons receive the same under both programs.
- anti-Leaky-bucket transfer: Program A is preferable compared to program B if there are two or more persons such that the poorest of them earns less in A than in B, and the total earnings in A are at least as much as in B. All other persons receive the same under both programs.

Policy preferences:

- early-life education: It is more important to direct funds to primary education rather than to increase the benefits of the poorest 5% of the people in society (assuming the expenditures needed are the same in the two cases).
- progressive income tax: It is better to fund the key activities of the public sector by an inheritance tax than an increase in the progression of the income tax (assuming the effect on government revenue is the same in the two cases).
- redistribution: A society should aim to equalize incomes.
- inheritance tax: A society should have an inheritance tax.
- capital gains tax: A society should have a capital gains tax.
- basic income: A society should have an unconditional basic income paid to every individual.

Survey items:

- rich deserve to be rich: High income earners in our society deserve their high incomes.
- poor deserve to be poor: Low income earners in our society deserve their low incomes.
- luck matters: It is mostly luck rather than effort that matters for how well an individual does economically in life.
- taxes affect effort: Changes in income taxes influence how much individuals choose to work.
- trust in party: There is a political party that can be trusted to implement the right income tax policies.
- effective administration: The public administration is very effective in limiting fraud, waste and abuse in the programs it administers.