

Non-selfish behavior: Are social preferences or social norms revealed in distribution decisions?*

Shaun P. Hargreaves Heap[†], Konstantinos Matakos[†], Nina Weber[†]

February 14, 2022

Abstract

People frequently choose to reduce own payoffs to help others. This non-selfish behavior is typically assumed to arise because people are motivated by social preferences. An alternative explanation is that they follow social norms. We test which of these two accounts can better explain subjects' decisions in a simple distribution game. Unlike previous studies, we elicit preferences and perceived norms directly for each subject. We find that norm-following explains people's distributive choices better than social preferences, and lack of confidence in one's social preference predicts norm-following. Our findings have implications for the strength of the Pareto criterion in welfare evaluations.

Keywords: social preferences, norms, distribution decisions, inequality, unselfishness, social identity, ambiguity, principles of justice, Pareto criterion, maximin

JEL Codes: A13, C90, D63, D64, D91, Z13

*We would like to thank the participants of seminars at the LSE, King's College London and the conference of the Economic Science Association for useful feedback. The data collection benefited from a grant from the Templeton Foundation and the ESRC.

[†]Department of Political Economy, King's College London, Bush House NE, London WC2B 4BG, UK. Correspondence to: nina.s.weber@kcl.ac.uk

I INTRODUCTION

People frequently behave non-selfishly. For example, people give to charities, make voluntary contributions to public goods and rich people vote for left-wing parties that will redistribute income to the poor (see [Alesina and Giuliano 2011](#); [Enke 2019](#); [Bregman 2020](#)). It is also one of the central insights from a range of experiments on decision making: e.g. in distribution decisions (see [Charness and Rabin 2002](#); [Fehr et al. 2006](#); [Bolton and Ockenfels 2006](#); [Cappelen et al. 2013](#)) and in public goods and prisoner dilemma decisions (see [Ledyard 1995](#), and the follow-up survey, [Chaudhuri 2011](#)). We address two important questions that arise from this evidence on unselfish behavior with an experiment.

First, do people behave unselfishly because they are motivated by social preferences or because they follow social norms? Second, is the character of unselfish behavior sensitive to the elicitation procedure?

The social preference answer to the first question accommodates unselfish behavior within the standard rational choice model in economics. It is sometimes given tautologically in the sense that behavior reveals a social preference when it is unselfish, but this is not the version of the social preference answer that interests us. We are concerned with whether people actually have social preferences they act upon: i.e. do they assess outcomes according to social preferences so as to make acting on social preferences an accurate psychological account of unselfishness? We ask the question in this form because there is an alternative explanation of unselfish behavior and it matters for welfare economics which accounts for such behavior.

The alternative is that people follow social norms: that is, people do what is generally regarded in society as the appropriate behavior in these circumstances. They are norm-followers, not preference satisfiers in these settings. While norm-following and social prefer-

ence satisfaction are sometimes used interchangeably in the literature (for instance, people might be said to have a social preference to follow a social norm (see e.g. [DellaVigna et al. 2020](#)), our goal is to make the two explanations both theoretically distinct and also capable of being distinguished in an empirical test. That is, we want, so to speak, to run a genuine horse race between social preferences and social norms, where the result matters. A third possibility, of course, is that people are selfish and are just making mistakes when they behave unselfishly.

The reason that the answer matters is because the Pareto criterion is the lynchpin of almost all welfare economics, but it can only apply to a world where people actually act so as to satisfy best their preferences (including social preferences if people have them). If people only acted ‘as if’ satisfying their preferences (or did not act to satisfy preferences at all), then it does not help when evaluating a policy to know that, in some other world where people did *actually* act to satisfy their preferences, this policy intervention would yield a Pareto improvement. The Pareto insight would apply to that other world where people do act to satisfy their preferences and not to the actual one where people only act ‘as if’ they were preference satisfiers or are not even ‘as if’ preference satisfiers at all.¹ Thus, to use the Pareto criterion for generating policy when people act unselfishly, we have to believe they are indeed doing so because they have a social preference. If people instead follow norms (as distinct from social preferences) when acting unselfishly, then the Pareto criterion is no

¹Suppose, for example, state A is Pareto superior to B because Simone’s preferences are better satisfied in A than B and no one is made worse off in A than B. The case for implementing A is clear because Simone is better off in A. If, however, Simone only acts ‘as if’ she had preferences that she sought to satisfy best and it is these ‘as if’ preferences that are better satisfied in A than B, then we have no way of knowing whether Simone is actually better-off in A than B because we no longer have an account of how Simone’s well-being is connected to outcomes in A and B. If she actually had these preferences, we would. That is why it is important to know whether people are actually preferences satisfiers. If she does not actually have these preferences, then we cannot judge whether A is better than B using the Pareto criterion and we need to develop some other framework for generating policy evaluations. More concretely and for the same reason, when cost-benefit analysis is used because it identifies potential Pareto improvements, the potential improvements have to be real and not ‘as if’ ones.

longer applicable and some other is required for making welfare judgments.²

The answer to the first question may also matter for positive economics, albeit more controversially and in a different way. Consistent behavior is all that is needed for prediction according to the revealed preference approach. Such consistency can, of course, be interpreted ‘as if’ people had preferences that they acted to satisfy best. But this is an optional ‘as if’ interpretation with the revealed preference approach. Although this revealed preference argument is widely accepted among economists, it is controversial among philosophers of social science who worry about the problem of induction (e.g. see [Hollis et al. 1994](#)). Their point is, that in the absence of a causal mechanism that explains why people behave in this consistent way, projecting predictively from previous instances of (consistent) behavior on to future ones relies on the principle of induction and this principle has only a circular or self-referential justification. It does not matter whether the psychological causal account involves preference satisfaction or norm-following for this purpose, but we must have reason to believe in one or the other before prediction based on consistency is causally warranted. We are not concerned here with evaluating the merits of this dispute in the philosophy of social science; we merely note that the explanation of why people behave unselfishly may be important not only for normative economics but also for causally warranted prediction in positive economics.

Our second research question arises because three different elicitation procedures are often used in the experimental literature: an individual makes a distribution decision as a member of the group knowing their position, as a member of the group behind a veil of ignorance regarding their position, and as an impartial spectator. We want to know whether the choice of elicitation mechanism makes a difference to the character of the unselfish behavior we observe and to the best psychological explanation of such unselfishness. This obviously

²Or in terms of the framework suggested by [Bernheim \(2009\)](#), the domain over which one can make such judgments shrinks.

matters whenever the specific experimental findings regarding unselfish behavior feed into and inform economic analysis (e.g. see [Durante et al. 2014](#)). We need to know for this purpose that the specific findings are not particular to the elicitation procedure.

To answer these questions, our experiment has subjects make three types of decisions regarding the distribution of income in a society. First, subjects are told how a particular distribution of income in a group arose and they then select a principle of justice, from a set of four, that they think should govern distribution for that group. The idea is that if a person does assess distributional outcomes through a social preference, then these preferences will be underpinned in this context by some principle of justice. This is a key connection for our argument. We justify it in part because this connection is made by economists when they categorize the social preferences that are revealed in distribution experiments: they use categories that relate to principles of justice (e.g. see [Charness and Rabin 2002](#); [Fehr et al. 2006](#); [Bolton and Ockenfels 2006](#); [Cappelen et al. 2013](#)). In addition, when political philosophers discuss what might inform moral behavior in such distribution decisions, it typically involves principles of justice (e.g. see [Rawls 1971](#)).³ Second, our subjects choose an actual distribution of income for this group from a set of possible distributions, each of which instantiates one of the four principles offered in the first decision. Finally we use the [Krupka and Weber \(2013\)](#) procedure for eliciting whether subjects perceive that there is a social norm regarding the distribution choice and what it is.

Our basic test of whether social preferences or norms (or neither) explain unselfish behavior

³It is also important that the distribution decision refers to a group of individuals. Had the decision referred to a dictator game, then it might be possible to argue that a more simpler fellow feeling, say of altruism, for another person underpinned the decision to give something to another person. It is more difficult to imagine how such a fellow feeling could explain such decisions when they affect a group of individuals. This is because the issue of how to weigh fellow feeling of this kind across the different individuals must arise in this context and this would seem to require, at least implicitly if not explicitly, a principle of justice to solve. We also make use of text analysis of comments made by the subjects at the end of the experiment when asked to explain the rationale for their decisions. From this it is plain that the currency of their offered explanations is shared or taken from that of our principles of justice.

is simple: does either the subjects' selected principle of justice and/or their perception of the social norm (or neither) predict their actual distribution decision. We further test for the influence of the elicitation procedure by having three treatments where we change the subject's relation to the group for whom these decisions are being made and we examine whether the relevant behaviors vary across them.

We are not the first to consider whether social norms might influence behavior (e.g. see [Krupka and Weber 2013](#); [Kimbrough and Vostroknutov 2016](#)) or whether such norms might explain behavior better than social preferences (e.g. see [Ellingsen et al. 2012](#); [Gächter et al. 2013](#); [Guala et al. 2013](#)). But our contribution is distinctive in two important ways. Our test occurs in a context where the norm following and preference satisfying explanations are genuine competitors: i.e. the difference is not merely semantic and nor are they potentially complementary accounts. Our test is also more direct. Both features are made possible because we use a distribution decision in the experiment.

The evidence from these earlier studies pitting social preference against social norms, in contrast, typically comes from trust and public goods games; and is mixed in its conclusions. This evidence is indirect in the sense that it usually depends on a particular theory of norm following and an assumption that the social preferences, if they exist, are stable across decision problems. With these assumptions, these earlier studies examine whether social preferences or norm following best organizes the data from trust and public goods games. Our approach is more direct (and requires fewer background assumptions) because we ask our subjects to identify through the first decision what, in effect, if they were motivated by a social preference, would be its character. This approach would be more difficult to do in trust and public goods games because there are a large number and variety of potential social or moral motives that might be in play as compared with the compact list of principles of justice that we use to identify possible social preferences in the distribution decision.

The other key difference is that these earlier ‘social preference versus social norm’ studies typically construe norm following as a coordination device when there are multiple Nash equilibria. In this way norm-following is a complement, or aid to preference satisfying behavior, rather than a challenge. However, while this is one way that norms might guide behavior, it is not the only one in the wider social sciences. There are more radical models of norm-following, more sociological or anthropological in origin but nevertheless still with an economic pedigree, that are a challenge rather than complement to the preference satisfying model. We are able to test these more radical senses of norm-following because our distribution decision is non-interactive and so there is no scope for norms to act as a coordinating device in an interaction that has multiple Nash equilibria. If norms guide behavior in our distribution decision, then it is potentially a much more fundamental finding.

On the first question, we find that norm-following is more important in explaining our subjects’ distribution choices than their social preferences. The descriptive aggregate data is powerfully suggestive in this respect. We find that people adhere to a variety of distributive principles. The most common principle is a form of ‘Meritocracy’ in our sample (around 38%) and the least common is Rawls’s Maximin principle (c.14%). In marked contrast, our subjects’ perception of the prevailing norm is concentrated on the Maximin distribution (49%); and so, critically for our conclusion, are the actual distribution decisions (50% are for the Maximin outcome).

In light of this somewhat surprising finding, at least for economists brought up on the preference satisfying model, we conduct several robustness checks. First, we run a second experiment that inverts the second and third decisions above so as to avoid the possibility that the actual distribution choices influence perceptions of the social norms. This further experiment also allows us to explore the origins of such norm-guided behavior. The second

experiment again reveals the primacy of social norms and it reinforces the social norm account by yielding some plausible insights into why people follow social norms. In particular, when subjects are confident in their choice of justice principle, they are more likely to follow it in the distribution decision and an individual's strong social identification helps build such confidence. Lower levels of confidence, in contrast, are more likely to lead to selfish or norm following behavior and this is consistent with Adam Smith's account of norm following in the *Theory of Moral Sentiments*. In our second and third robustness check, we examine with further surveys two additional possibilities that might have contributed to the weak evidence in favor of social preferences. One is whether people are guided by more than one principle of justice and if such secondary principles might explain the drift to maximin outcomes. The other is whether an inability to associate a principle of justice with a distribution outcome in the second decision might weaken the evidence of social preferences in those decisions. Again, we conclude the original result favoring norm-following is robust to these considerations.⁴

On the second question, we find that none of the expected variation across the different elicitation procedures appear in our data. While not expected at the outset, this is not so surprising given our first finding. The differences and debates around the choice of elicitation procedure are typically premised on the idea that people are preference satisfiers. This is why the different mechanisms seem likely to produce different results because they either do not or do allow, but in different ways, selfish preferences into decision making as well as social ones; and this is why there is a debate over which should be used. However, if decision making is mainly based on norm-following, then there is no reason to expect to find these preference-satisfying based differences across the elicitation procedures. This is what we find: the character of unselfish behavior and its apparent explanation does not materially depend on the elicitation procedure.

⁴We also subject these results to various robustness checks regarding the wording of the principles, see later footnote 7 and appendix section C.7.

Our main contribution, then, is to test whether social preferences or norm-following best explains unselfish behavior in a setting where norm-following supplies a distinct alternative model of behavior to that of preference satisfaction. Our findings are in favour of norm-following. This has important implications. Our experiment suggests that the use of the Pareto principle in welfare economics has, in general, a weak foundation because whenever people behave unselfishly such behavior is not well captured by a preference satisfying model. In particular, it cannot be assumed that unselfish behavior reveals social preferences that can then be entered into a social welfare function for the purposes of developing policy recommendations. Our robustness checks reinforce this general conclusion but also suggest that it is possibly less of a problem in societies where individuals have a strong sense of social identification.⁵

In the next section, we review the background literature on which we draw to develop our hypotheses. Section III sets out the experimental design and Section IV gives the results. Section V briefly introduces the second robustness check experiment. We discuss the results and conclude the paper in Section VI.

II BACKGROUND LITERATURE AND HYPOTHESES

We have two research questions and two sets of hypotheses which we elaborate below.

⁵In so far as more homogenous societies engender social identification (Alesina and Glaeser 2004), then this result leads to the prediction that social preferences are more likely to guide unselfish behavior in homogenous societies than in heterogeneous ones, where social norms are more likely to explain such behaviors.

A. Social preferences versus social norms hypotheses

When people act non-selfishly, the rational choice model offers a simple explanation: people have ‘social’ as well as ‘selfish’ preferences. We call this the social preference hypothesis (H1). People care not just about how their interests are affected but also how others fare in any outcome. The rational choice model is usefully quiet on the character of preferences and so the incorporation of social preferences presents no threat to the model. All that matters is that behavior should be consistent in a manner that is representable by a preference ordering (e.g. see [Andreoni and Miller 2002](#)).

H1: Social preferences predict the character of unselfish behavior.

An alternative possible explanation is that people behave unselfishly because they are guided by a social norm. This can be variously understood. It is a traditionally more sociological way of explaining behavior (e.g. see [Parsons et al. 1949](#); [Durkheim 2013](#)) and if understood literally it can attract the charge of turning people into cultural or social dopes. To avoid this charge and retain plausibility, individuals have actively to participate in the decision making process in some way (at least at some times). There are several ways in which this has been imagined while allowing for the influence of norm-guided behavior and we distinguish between those that complement and those that challenge the preference satisfying model of behavior.

Those that complement the preference satisfying model either introduce, as just discussed, norms as an informational devices that aid equilibrium selection in games with multiple Nash equilibria (see also [Binmore 2010](#)), or they allow that norms might help constitute the social preferences which people act upon. In both cases, individuals still make decisions by acting so as to satisfy best their preferences. We preclude the former by design because

there is no interactive decision making in our experiment. We focus, therefore, on the latter form of complementarity. [Duesenberry et al. \(1960\)](#) famously illustrates the idea that norms help constitute preferences and this idea has recently received increasing attention as result of the introduction of social identification theory into economics (see, respectively, [Tajfel et al. 1979](#), [Akerlof and Kranton 2000](#)).⁶ In social identification theory, it is argued that people gain a sense of identity by behaving in a way that corresponds to the norms of their group. This gives them a sense of identity because their group’s norms differ from those of other groups. Thus to act in accord with the norm is to create a new reason for acting in that way: an identity. [Akerlof and Kranton \(2000\)](#) represent this idea through a ‘new’ argument in a person’s utility function. Thus the norms of one’s group help constitute a person’s preferences (i.e. their utility function), but they still act so as to satisfy best their preferences (maximize their utility). For such individuals, their social preferences and their group’s norms are essentially one of the same, at least for those who identify strongly with their group. We call this the norms as social preferences hypothesis, H2; and although it allows a role for social norms, it effectively makes the competition between social preference and social norm redundant because they are one of the same.

H2: Social norms constitute social preferences and so both predict character of unselfish behavior.

In contrast, there are two ways in which being guided by a norm both involves individual volition and also marks a departure from preference satisfying behavior. Both turn on a different epistemic predicament: an existential one. Individuals face this predicament when they do not have well defined preferences to act upon; and so they turn to norms as a source of information/guide on what to do.

⁶[Bicchieri \(2005\)](#) and [Gintis \(2010\)](#) in different ways straddle this distinction between the two complementary routes by having norms both help constitute player utility functions and play a coordinating role. See [Paternotte and Grose \(2013\)](#) for a review of these differences.

In one case, individuals do not have a relevant preference. For instance, the outcomes associated with an individual's possible actions might be so novel (because they involve some new people, or products, or old ones in new situations) that individuals cannot evaluate them; and in these circumstances, they treat other people's behavior as social information regarding how to value them. They take their cue, in other words, for what is valuable from what others do and so conform to their behavior. This type of conformism may have evolutionary as well as sociological origins (see [Apesteguia et al. 2007](#), and [Alger and Weibull 2013](#)) and there is some experimental evidence in its support (see [Fatas et al. 2018](#)). We call this the norm as conformism hypothesis (H3a) and such norm guided behavior is distinct from preference satisfying behavior because it arises when individuals do not have the relevant preferences to guide them. We note in passing that such norm guided behavior need not always relate to unselfish behavior but is more likely to in settings involving other people.

The second version of this existential epistemic predicament has an eminent economics pedigree: it is set out by Adam Smith in his *Theory of Moral Sentiments*. People in this instance have preferences (unlike the above) but face a problem of acting in good faith upon them when social preferences conflict with what selfishness commends. The problem arises in this case when the interpretation of what is required by a social preference involves some discretion and, when the social preference is in some degree opposed to a person's selfish preference, a person will know that their own interpretation of the social preference could be self-serving (i.e. a 'bad' faith interpretation). To avoid this suspicion and so experience a genuine or authentic pleasure of satisfying in some degree one's social preferences (in 'good' faith, as it were), there has to be some standard external to the individual for the interpretation of how to act on a social preference authentically. This is what social norms supply and why they are followed. This is how Adam Smith makes the same point.

“The opinion which we entertain of our own character depends entirely on our

judgments concerning our past conduct. It is so disagreeable to think ill of ourselves, that we often purposely turn away our view from those circumstances which might render that judgment unfavourable. He is a bold surgeon, they say, whose hand does not tremble when he performs an operation upon his own person. . . . So partial are the views of mankind with regard to the propriety of their own conduct. . . . Nature, however, has not left this weakness, which is of so much importance, altogether without a remedy; nor has she abandoned us entirely to the delusions of self-love. Our continual observations upon the conduct of others, insensibly lead us to form to ourselves certain general rules concerning what is fit and proper either to be done or to be avoided” (Smith 1759, Part III, ch iv).

We call this the norms as ‘good’ faith or authenticity devices hypothesis (H3b) and it has the same implication as H3a regarding what predicts the character of unselfish behavior, hence H3 below covers both H3a and H3b. We note that with H3b a norm is used to accommodate social as well as selfish preferences. Thus, H3b norm guided behavior is again distinct from individual preference satisfying behavior both because it is an accommodation with their selfishness and, importantly, because the person’s social preference need not be the same as those held by others in society and it is the latter that actually provides the normative guide to action.

H3: Social norms predict the character of unselfish behavior.

It may be tempting to imagine that these existential, epistemic based, norm guided behaviors can nevertheless still be subsumed under the preference satisfying model by allowing for individuals to have a preference, say, respectively, for conforming to a social norm and/or authenticity. Thus, it is these authenticity/conformity preferences that explain why behavior is guided by the relevant norms and so there need be no break with the preference satisfying model of action to cover these kinds of behaviors. The difficulty with this strategy is that it stretches what is an elastic concept of preference satisfaction too much. For instance, a preference for conformism in these circumstances amounts to having a preference for creating a preference by simply following what others do. At best this is following other people’s

preferences and not your own and, when all do this, behavior has no anchor in anyone's preferences at all. Each is simply following what others do.

Likewise, acting on a preference for authenticity in the Adam Smith version of norm guided behavior creates a similar problem. A preference for authenticity is, in effect, in these circumstances a preference not to be guided by one's own preferences. This is self-contradictory in a way that threatens to make the idea of acting on preferences meaningless. Preference satisfaction has to be a falsifiable if it is to be a meaningful concept and so there must be some limit to the possible preferences that can be added to the model so as to account for behavior. Otherwise whenever a behavior occurs that cannot be understood from the existing list of preferences, one can simply add a new preference for that behavior (whatever it is) and the model becomes effectively unfalsifiable. To be falsifiable there has to be some constraint on this type of addition to the list of preferences: there must be limits on what might count as a preference and a natural candidate for exclusion from such a list of possible preferences, is a 'preference not to act on one's preferences' because it involves an internal contradiction.

B. Identification of social preferences through principles of justice

We have already argued that the choice of a distribution decision for a group of individuals, unlike public goods and trust games, usefully constrains the moral foundations for pro-social behavior to principles of justice. This is what has been assumed by experimentalists in the past as they categorize social preferences and it is what is suggested in political philosophy. We therefore ask our subjects, as a method for revealing the character of their social preferences, to select a principle of justice that they believe ought to apply to a group of people. Our choice of principles of justice for this purpose comes from that practice among experimentalists and the discussions in political philosophy. On this basis we identify 4 broad principles.

The first has its origins in Marxian political philosophy. Marx famously proposed that, ideally, distribution would follow the dictum ‘from each according to his ability, to each according to his need’. In the absence of knowledge about differences in need, this translates into a familiar left-political preference for equality; or, to put this round the other way, an aversion to inequality. An aversion for inequality has been formulated by [Fehr and Schmidt \(1999\)](#) and [Bolton and Ockenfels \(2000\)](#) and there is considerable experimental evidence that is consistent with such an aversion guiding in various degrees individual distribution decisions (e.g. see [Charness and Rabin 2002](#), [Fehr et al. 2006](#), and [Bolton and Ockenfels 2006](#)). We represent this principle with the following statement in the experiment.

- Inequalities should be minimized.

The second principle comes from [Rawls \(1971\)](#). His second principle of justice is the so-called ‘difference principle’ and it recommends that once equal freedoms have been guaranteed (the first principle), we should prefer societies that produce the best outcome for those who are worst off: i.e. the Maximin principle. Given the central place of Rawls in liberal political theory, this is an obvious candidate principle. However, it is worth noting that while there is some experimental evidence that is consistent with Maximin preferences over distribution decisions (see [Charness and Rabin 2002](#), [Engelmann and Strobel 2004](#), and [Fisman et al. 2020](#)), the evidence is probably not as strong as that of inequality aversion (e.g see [Fehr et al. 2006](#)). The statement of this principle in the experiment is:

- Inequalities are only justifiable if they improve the position of the least well-off group in society.

The third principle is from the philosophy of utilitarianism and the suggestion that societies should be arranged to produce the ‘greatest happiness for the greatest number’. In the absence of specific knowledge about how income translates into happiness for different people, this becomes a preference for arrangements that produce the highest average income level. This, for example, is the implication of Harsanyi’s ([1980](#)) derivation of utilitarianism

from the same veil of ignorance procedure as Rawls when individuals are expected utility maximisers (and not deciding using maximin). The arrangement that produces the highest average income is also associated with exhausting all potential Pareto improvements (when allowing for compensation schemes) and so reflects a concern for efficiency. There is again considerable experimental evidence that is consistent with such efficiency or utilitarian preferences explaining behavior in distribution decisions (especially among economics students, see [Engelmann and Strobel 2004](#), and [Fehr et al. 2006](#)). The statement used for this principle in the experiment is:

- Income should be distributed to maximize the average income in society.

Our final principle is meritocratic: that is, people should be rewarded according to their ability and talents. This is a version of a desert theory of justice and in our particular context it is also what Nozick’s libertarian political philosophy ([Nozick 1974](#)) would commend: i.e. that we respect the outcomes that come from the free exercise of individual choice. Again there is experimental evidence that distributional choices are in part guided by a meritocratic concern. For example, [Cappelen et al. \(2013\)](#) find that people are less inclined to redistribute when the inequalities emerge from individual choices than when they emerge as a matter of luck. Meritocracy is phrased:

- Individual income should be based exclusively on his/her ability and talents.

We conduct a post-experiment check, that we comment on later in section [V](#), on whether these types of ideas are used by our subjects to explain how they selected a distribution outcome in an open commentary box at the end of the experiment.

C. Elicitation mechanisms hypotheses

Three elicitation mechanisms for the revelation of social preferences have often been used in the literature (e.g. see [Durante et al. 2014](#) who, like us, uses versions of all three). The debate over which is to be preferred is typically premised on the social preference model

of behavior (i.e. H1 and/or H2 holds). This is because the elicitation mechanisms differ according to whether or how selfish preferences also enter into decision making. Thus, they may or may not reveal social preferences or some combination of social and selfish ones. It is these differences that are the basis for hypotheses H4 and H5. In this way, since the hypotheses are premised on the social preference interpretation of unselfish behavior, in so far as we do not find support for them, then this may also be taken as evidence against the social preference model.

The first mechanism, the Impartial Spectator, used notably by [Cappelen et al. \(2013\)](#), so distances any selfish preferences from an individual's distribution decision that the decision can only reveal their social preferences. Subjects are asked to make decisions for a group of people and the decisions will not affect the subjects' own pay-offs.

In the second procedure, subjects make the decisions for a group of people that, this time, they belong to, but behind a Veil of Ignorance. As they belong to this group, they will be affected by the distribution decision, but they do not know when making it what position they will occupy under any particular distributional choice. As compared with the Impartial Spectator procedure, this gives the subjects a stake, if an obscure or uncertain one, in the outcomes of a decision. There are two possible ways of interpreting the decisions made with this procedure. One is that they reveal a person's social preferences because selfish interests are rendered obscure by the Veil of Ignorance. The other, and this is what [Rawls \(1971\)](#) originally argued, holds that the procedure reveals what justice requires and that those with 'moral' personalities will then be guided by this. In other words, the mechanism does not reveal a person's social preferences, it shows how to act justly in a society where selfish and other-regarding interests are not aligned. Thus, the Rawlsian procedure, on this view, recommends a rule to guide action. This is a rule form of rationality and is actually one step towards the norm-following model of action. If, in addition, for example, such a rule is

shared for the epistemic reasons suggested by Adam Smith, then it would, in effect, be like Adam Smith's version of norm-following behavior.

Independently of which interpretation of the veil of ignorance is to be preferred, it has been argued that this procedure will deliver a particular substantive distribution decision. Rawls argues that individuals facing this uncertainty will use the Maximin rule and so select the Maximin distribution. [Harsanyi \(1955, 1980\)](#), in contrast, argues that individuals face the uncertainty as expected utility maximisers and so choose the Utilitarian/Efficient distribution. Thus, we expect the veil of ignorance procedure to skew our subjects' distribution decisions to one or other of these outcomes. Since there is no reason to suppose that Impartial Spectator's social preferences are exclusively developed through the Veil procedure, we do not expect to find that our subjects, when acting as impartial spectators, will be similarly skewed towards these two distribution outcomes. H4 follows.

H4: Individual decisions in Veil of Ignorance distributions decisions will be skewed towards either Maximin or Utilitarian/Efficiency as compared with Impartial Spectator.

Our third elicitation procedure removes the Veil of Ignorance. Subjects belong to the group and know what position they will occupy in the income distribution when they decide on both the principle and the distribution. Thus individual distribution decisions reveal some combination of selfish and social preferences. H5 follows and again it is premised on H1.

H5: Selfishness helps predict the distribution decisions with the Non-Veil of Ignorance procedure.

The combination of motives under this procedure is unfortunate if the purpose is to discover the character of social preferences alone, but the procedure has the advantage of incentiviz-

ing subjects clearly. It may be attractive for this reason and so it becomes important to know whether the contamination for introducing selfish motives is significant (i.e. whether H5 holds).

Our final elicitation hypothesis relates to whether the procedure may affect the propensity to reveal behavior that is more or less consistent with the social preference or norm-following account of unselfish behavior. Since the elicitation mechanism is a known context from the outset of the experiment for both the principle of justice decision and distribution decision, there is no obvious reason for supposing that we will observe any difference in the frequency of principle-distribution consistency across the elucidation mechanism treatments. For example, in the non-Veil of Ignorance mechanism, selfishness is as likely to be a consideration in the choice of principle as in the distribution decision, thus in so far as selfish and social preferences explain decisions then we expect principle-distribution consistency.

The efficacy of the [Krupka and Weber \(2013\)](#) procedure should likewise be the same across elicitation mechanisms, but it is possible that norm-following in the distribution decision might vary. For instance, since the Veil of Ignorance can be interpreted as a rule generating device, it might encourage rule following and precisely because the non-Veil of Ignorance requires an individual to consider how to combine selfish and social preferences, it may too encourage thinking in terms of rules for Adam Smith-like reasons in ways that the Impartial Spectator need not. However, in neither of these cases does the mechanism encourage the thought that rules need be shared and so become norms. Thus we see no obvious reason why the distribution-norm consistency should vary across the elicitation procedures. H6 follows.

H6: The frequency of social preference-distribution consistency and perceived social norm-distribution consistency does not differ across elucidation procedures.

III EXPERIMENT DETAILS

Subjects are told, by way of background, that a group of people are asked to do a quiz and their answers generate income. Their performance is ranked from the bottom 20% of performers to the top 20% in Table 1, where we give the average income generated for a person in each 20% performance band.

Table 1: Average Income per Quintile

<i>Performance Level</i>	<i>Average Income</i>
Bottom 20% of performers	£20
2nd 20%	£30
3rd 20%	£40
4th 20%	£70
5th 20%	£110

Decision 1: Choice of Principle.

Subjects are asked which of the following statements best describes how they think income should be distributed in this group of people. The statements have already been given and are randomly ordered in the experiment.⁷

It is important that they are asked about which principle best describes how income should be distributed in this group after they know the status quo distribution and how it arose. This is because the attractiveness of a principle may depend on the situation to which it might be applied. For instance, even those who are averse to inequality may not be so concerned to minimize income differences when they are already small; and another principle may become more important. This decision together with Decision 3 allow us to test H1.

⁷We report the results of a robustness check using an alternative wording for the maximin and inequality aversion principles in section C.7. of the appendix. Our results are robust to this test. The exact wording of the alternative statements can be found in appendix section A.3.8.

Decision 2: Elicitation of Social Norm regarding Principle.

All the participants of the study are now asked to select a principle from this list above and they are told that ‘you will be rewarded with a bonus payment of 50p if you select the principle chosen by most of the participants’. This is similar to a beauty contest with multiple equilibria where no one equilibrium is favored. If subjects choose a particular principle, it follows that this principle is the perceived social norm, as there is no strictly rational reason to choose one over the others. This is the [Krupka and Weber \(2013\)](#) mechanism for eliciting social norms. The only difference is that we apply this to the same population of subjects who make Decision 1; whereas Krupka and Weber use another subject pool.⁸ The specific purpose of this aspect of the design is to allow the test of H2 by comparing the answers with those in Decision 1.

Decision 3: Choice of Distribution.

Subjects are now informed that the income generated by the quiz in this group of people can be distributed in 4 possible ways and they are asked to decide on the distribution. This decision is incentivized in two of the three elicitation treatments because the subjects know that it, together with their likely quiz performance, will affect their final payoff. The options are given in [Table 2](#) for the income level for each person in each quintile, and again the order is randomly generated. We also give the total for a representative sample of 5 individuals, one from each quintile, to bring out that one is more efficient.⁹

The first distribution yields the smallest average difference between incomes and we associate

⁸For our purpose, it is more sensible to elicit the social norms from the same subject pool as make the distribution decision since norms may vary with subject pools and we wish to know whether our subjects are guided if at all by their perceived social norms and not some other group’s. In the robustness check experiment (Robustness Check 3) where we invert decisions 3 and 4, we find a similar distribution of perceived social norms, but in a further subject pool in Robustness Check 2 in the online appendix A.3.4., there are some differences.

⁹We also run a robustness check where we report the average income, rather than the total income, for each distribution option. The results can be found in online appendix, C.2. The different wording has no effect on subjects’ choices.

Table 2: Distribution Options

<i>Performance Level</i>	<i>Inequality Aversion</i>	<i>Average Income</i>		
		<i>Maximin</i>	<i>Meritocracy</i>	<i>Utilitarianism</i>
Bottom 20%	£30	£40	£20	£20
2nd 20%	£60	£40	£30	£30
3rd 20%	£60	£50	£40	£50
4th 20%	£60	£60	£70	£70
5th 20%	£60	£80	£110	£110
Total	£270	£270	£270	£280

Notes: The exact wording and presentation of the distributions to respondents can be found in online appendix E.

this outcome with the version of the Inequality Aversion principle that says inequality should be minimized. The second distribution is more unequal in the sense that it has a higher average difference in incomes, but it has a higher average income for the lowest quintile. This is the Maximin outcome. The third distribution is the one based on quiz performance and we associate this outcome with the Meritocratic principle of rewarding according to ability and talent. The fourth distribution is more efficient because it is the same as the initial quiz distribution except that the middle quintile earn £10 more. So we associate this with the Utilitarian principle of maximizing the average income. Of course, the distributions are not actually labeled with their corresponding principle in the experiment.

Decision 4: Elicitation of Social Norm regarding Distribution.

This replicates Decision 2 but is now directed at the actual distribution (and not the principle): i.e. ‘you will be rewarded with a bonus payment of 50p if you select the distribution chosen by most of the participants.’ The purpose of this decision is to identify a possible social norm regarding distribution decisions that is indeed distinct from social preferences when testing H3.¹⁰

¹⁰We also run a robustness check where we ask respondents to make decisions 3 & 4 prior to decisions 1 & 2 to test whether the order affects choices, preference following, or norm following. The results are reported

Treatments.

The treatments are distinguished by the relationship that the subjects have to the group of people that has done the quiz and for whom the subjects are making decisions.

In the Impartial Spectator (Treatment 1), the subjects are asked to decide on the principle and distribution for that group and they are explicitly told that they are not a part of this group.

In the Veil of Ignorance (Treatment 2), the subjects are told they belong to this group ('you will participate in the above mentioned quiz' and this will 'affect your bonus payment', you decide for 'your group'), but they do not know their quintile position or what the quiz consists of. They do know, however, that they will do a version of the quiz later and that their choices in the distribution decision together with the quintile position that comes from how well they did on the quiz will affect their final payment.

In the non-Veil of Ignorance (Treatment 3), the subjects are told in the same way as Treatment 2 that they belong to the group of people doing the quiz, but, in contrast to Treatment 2 and before they make any decisions, they are, in addition, asked to answer a sample of quiz questions. Their answers are used to give the subject a prediction of their likely quintile position in the actual quiz. So subjects in Treatment 3 both know they are making decisions for their group and their likely own actual income under each distribution.

The experiment was conducted online in November and December of 2019 using Prolific Academic. There were 2,408 subjects from the UK, US and Europe and they earned on

in section B.1. and C.8. of the appendix. The reversed order does not affect our main results or the choices subjects make in any noticeable way.

average £1.55. The participation time was on average 8 minutes and 17 seconds.¹¹

IV RESULTS

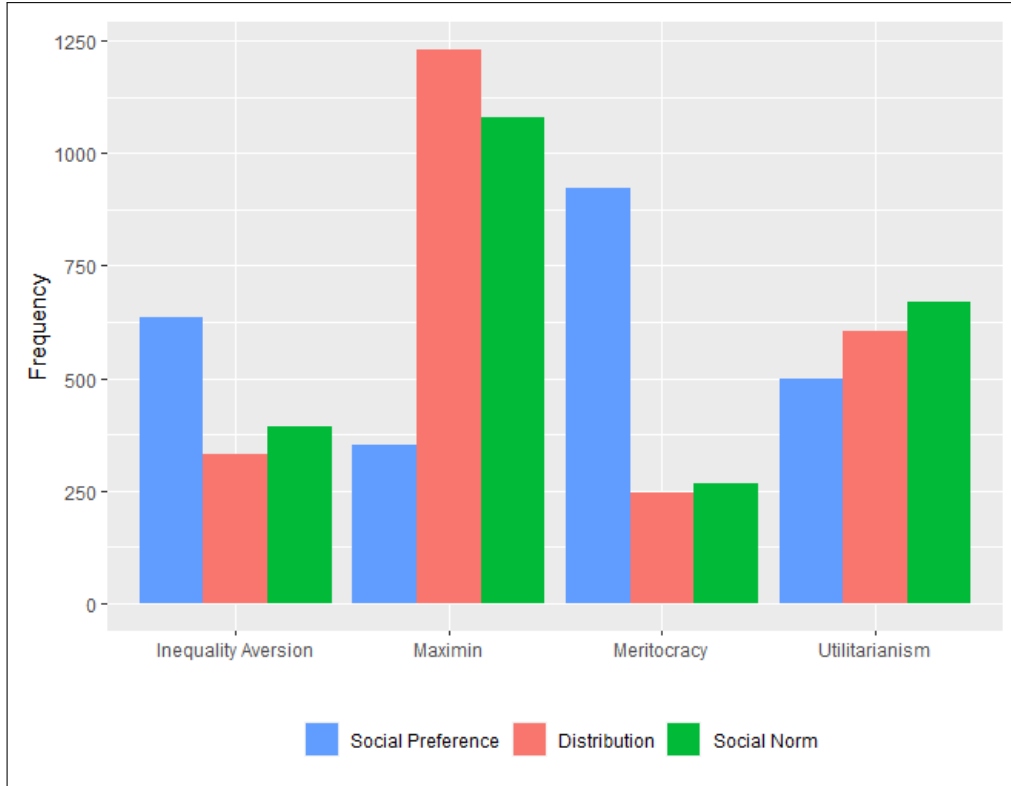
We initially focus on the horse race between social preferences (H1) and distribution social norms (H3) as predictors of distribution decisions, then turn later to H2 on whether the influence of social preferences is effectively the same as that of social norms because the one helps constitute the other.

Figure 1 gives the frequency of principle choices (decision 1), distribution choices (decision 3) and perceived distribution social norms (decision 4), in the aggregate for all three treatments. It is apparent that principle choice is a poor predictor of the distribution choices in the aggregate as compared with perceived social norm: i.e. blue (principle) and red (distribution) bars are poorly correlated whereas red (distribution) and green (social norm) bars are highly correlated. The only principle that predicts well the choice of distribution in the aggregate is Utilitarianism/efficiency. Those who subscribe to Meritocracy and Inequality Aversion as principles shift in the aggregate to the Maximin distribution and the latter is also there the most frequently perceived social norm.

Table 3 shows the aggregate data in a different way. It plots the frequency with which individuals who select a particular principle or identify a particular perceived distribution norm actually choose among the distribution options. The drift to the Maximin distribution for each chosen principle is evident in the first part of the table. For all subjects, except for those who chose the meritocratic principle, maximin is the most frequently chosen distribution. Indeed, except for those who chose the maximin principle, the principle choice only predicts the distribution choices of 16-19% of subjects. For comparison, suppose a person chose their preferred principle and then randomly selected the distribution: i.e. the principle choice has

¹¹Details of the sample composition and individual waves of the experiment can be found in online appendix A.

Figure 1: Frequency distribution of distribution choice, norm and preference



no influence on the distribution decision. It follows holding principle 'x' would nevertheless 'correctly' predict distribution choices 25% of the time with such random behavior. This means that for people who choose the Inequality Aversion, Meritocracy and Utilitarian principles in our experiment, their actual chosen distribution outcome are no better predicted than they would be had those distribution outcome decisions actually been random. The same comparison of the congruence between perceived social norm and distribution decision is stark. For all subjects in Table 3, social norms can explain a significantly larger percentage of distribution choices. Even for those who did not choose maximin, social norms explain the distribution choices of 35-53% of subjects—significantly better than would be the case if subsequent decisions were random.¹²

This contrasting assessment of the aggregate data is reinforced by simple correlation co-

¹²Indeed, for each social norm the proportion of consistent actual distribution choices would have been very unlikely to have arisen by chance had distribution choices been random ($p=0.000$).

efficients between distribution choices and principles (= -0.87) and between distribution choices and distribution norms (= 0.99).

Table 3: Social Preference and Norm by chosen Distribution

<i>Social Preference</i>	Chosen Distribution			
	Inequality Aversion	Maximin	Meritocracy	Utilitarianism
Inequality Aversion	18.90%	62.52%	5.83%	12.76%
Maximin	11.43%	68.86%	6.00%	13.71%
Meritocracy	6.60%	34.42%	16.56%	42.42%
Utilitarianism	21.84%	55.11%	6.61%	16.43%
<i>Social Norm</i>				
Inequality Aversion	47.46%	33.50%	5.58%	13.45%
Maximin	8.33%	76.60%	3.79%	11.29%
Meritocracy	9.06%	29.06%	35.09%	26.79%
Utilitarianism	4.34%	29.04%	13.17%	53.44%

We turn now to the individual level evidence. Table 4 reports, for each choice of a particular distribution, whether it helps in predicting that choice to know that a person’s chosen principle was consistent with that choice and whether that person’s perceived social norm was consistent with that choice, plus a series of other possible explanatory variables. Thus in the first column, the dependent variable is a dummy taking a value 1 when the individual chose the inequality averse distribution (otherwise 0), the first explanatory variable (social preferences) is a dummy variable taking value of 1 when that individual chose the inequality averse principle (0 otherwise) and the second explanatory variable (social norm) is a dummy taking a value 1 when that person’s perceived social norm is inequality averse (and 0 otherwise). The next two variables are treatment dummies and there follow a series of individual controls: e.g. gender, age, income etc.

For each distribution choice, it helps in predicting that choice to know whether the person chose the associated principle and whether that person perceived the social norm to be

that distribution. The coefficients on the social preference and social norm variables are always positive and significant. The only exception is the social preference coefficient in the utilitarian distribution choices where the coefficient is negative. In other words, if a person chose principle X, it is more likely that they chose distribution X, except when X is Utilitarianism, and if a person perceives the social norm as X then they also choose distribution X, for all X. Crucially, though, the size of the coefficient on social norms is always significantly larger than that on social preference. Taking Meritocracy as an example, subjects who have a social preference for meritocracy are 2.6 times as likely to choose the meritocratic distribution compared to everyone else, while subjects who perceive the social norm to be meritocratic are 7.3 times as likely to choose the meritocratic distribution.

The second part of Table 4 focuses on the non-Veil of Ignorance treatment. We can introduce a potential selfishness explanatory variable in this treatment because subjects answer some sample questions from the quiz and get a prediction of their likely quintile position. This is a dummy taking the value of 1 when the distribution choice also accords with a person's selfish interest (maximizes own expected material returns); 0 otherwise. The coefficients on the social preference and social norm variables are very similar to those models including all treatments, so the introduction of a possible selfish motivation does not seem to materially affect the influence of social preference and social norms on choices.

Table 4: Logistic regressions of distributive choices for all treatments

	All Treatments				Non-Veil of Ignorance Treatment			
	Inequality Aversion	Choice of Distribution		Utilitarianism	Inequality Aversion	Choice of Distribution		Utilitarianism
		Maximin	Meritocracy			Maximin	Meritocracy	
Social Preference	0.519*** (0.152)	0.806*** (0.145)	0.960*** (0.154)	-0.648*** (0.150)	0.465* (0.260)	0.538** (0.260)	0.542* (0.282)	-0.115 (0.237)
Social Norm	2.515*** (0.142)	2.086*** (0.101)	1.985*** (0.167)	2.052*** (0.112)	2.439*** (0.243)	2.275*** (0.183)	1.637*** (0.296)	2.020*** (0.187)
Selfishness					0.232 (0.236)	-0.486*** (0.179)	0.559** (0.252)	0.062 (0.187)
Treatments								
<i>Veil of Ignorance</i>	0.612*** (0.175)	-0.061 (0.120)	0.006 (0.179)	-0.197 (0.138)				
<i>Non-Veil of Ignorance</i>	0.536*** (0.178)	-0.156 (0.120)	-0.081 (0.185)	-0.009 (0.137)				
Sample								
<i>United Kingdom</i>	-0.068 (0.185)	0.061 (0.132)	0.350 (0.219)	-0.127 (0.144)	-0.231 (0.327)	0.211 (0.247)	-0.050 (0.368)	-0.002 (0.238)
<i>United States</i>	-0.189 (0.247)	-0.085 (0.166)	0.518** (0.248)	-0.024 (0.175)	-0.326 (0.427)	-0.132 (0.307)	0.850** (0.403)	-0.117 (0.300)
Quiz Performance	-0.080 (0.068)	0.037 (0.047)	-0.130* (0.077)	0.037 (0.055)	0.004 (0.107)	-0.085 (0.083)	0.073 (0.134)	0.041 (0.095)
Income	0.020 (0.070)	-0.021 (0.046)	0.052 (0.066)	-0.010 (0.052)	0.059 (0.110)	-0.030 (0.080)	-0.079 (0.113)	0.069 (0.086)
Female	0.241 (0.155)	0.101 (0.109)	0.244 (0.173)	-0.364*** (0.121)	0.572** (0.269)	-0.292 (0.204)	0.691** (0.338)	-0.407* (0.207)
Left-Right	0.087 (0.095)	0.247*** (0.067)	-0.082 (0.100)	-0.290*** (0.074)	-0.113 (0.161)	0.378*** (0.118)	-0.055 (0.175)	-0.324*** (0.122)
Age	0.066 (0.065)	-0.016 (0.046)	0.016 (0.070)	-0.050 (0.052)	0.124 (0.119)	-0.089 (0.086)	0.269** (0.130)	-0.158 (0.097)
Risk seeking	0.082** (0.033)	-0.064** (0.023)	0.010 (0.033)	0.034 (0.026)	0.114* (0.058)	-0.118*** (0.041)	0.034 (0.056)	0.067 (0.046)
Student	-0.006 (0.199)	0.092 (0.138)	0.215 (0.202)	-0.184 (0.150)	0.145 (0.342)	0.120 (0.222)	0.826** (0.352)	-0.616** (0.261)
Economics	-0.180 (0.180)	-0.245* (0.129)	0.043 (0.189)	0.342** (0.137)	-0.386 (0.288)	-0.229 (0.222)	0.056 (0.316)	0.388* (0.228)
Constant	-4.029*** (0.552)	-1.705*** (0.380)	-3.018*** (0.598)	-0.349 (0.425)	-3.484*** (0.942)	-1.189* (0.650)	-4.583*** (1.154)	-0.354 (0.746)
Session Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	2,219	2,219	2,219	2,219	733	733	733	733
Pseudo R-squared	0.210	0.191	0.143	0.175	0.200	0.216	0.122	0.172

Notes: Estimates come from a logistic regression. Social Preference is a binary variable equal to 1 if the subject's choice of principle matched the distribution. Social Norm is a binary variable equal to 1 if the subject's social norm matched the distribution choice. Selfishness is a binary variable equal to 1 if the subject chose the distribution that maximises the payoff of the quintile they were placed in based on their example quiz answers. The reference category for the treatment variables is the Impartial Spectator treatment. The reference category for the sample variables is Western Europe. Quiz performance ranges from 0 to 5 depending on how many questions the subject answered correctly. A higher value on the left-right variable indicates a more left-wing orientation on economic policy. Risk preferences are self-reported on a scale from 0 to 10 with 10 being the most risk-seeking option. Student is a binary variable equal to 1 if the subject is currently studying towards a degree and Economics is a binary variable equal to 1 if the subject has ever studied a course on Economics at University. Robust standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

This means, to take up the latent third possible horse in the explanation race of unselfish behavior, that deviations from selfishness are not simply random errors: instead these deviations are predicted either by a person's social preference or their perceived social norm.

Results 1, 2 and 3 follow.

Result 1 (weakly supporting H1): There is individual level evidence that social preferences help predict distribution choices, except in the case of Utilitarian choices. But there is no evidence in the aggregate data that social principles help predict distribution choices.

Result 2 (supporting H3): Perceived distribution social norm predicts the choice of distribution in aggregate and individual level data.

Result 3 (supporting H3 over H1): Perceived distribution social norms have a stronger predictive effect on distribution choice than social preferences in the aggregate and individual level data.

Turning to H2, we consider whether there is evidence that any part of the influence of social preferences on distribution decision making may reflect that of social norms because justice principle social norms constitute social preferences. Recall we elicited two possible norms through Decisions 2 and 4: perceived norms regarding the principle of justice and perceived norms regarding the distribution decision. We have so far found that the perceived distribution norms predict actual distribution choices better than individual social preferences. This means perceived distribution norms are unlikely to predict individual social preferences well. Indeed only 24% of our subjects chose a principle that is the same as their perceived distribution norm. To put this in perspective, had both decisions been random we should expect this coincidence with a 25% frequency (i.e. the frequency of coincidence we find is no different from what would arise randomly when each decision is independent of the other).

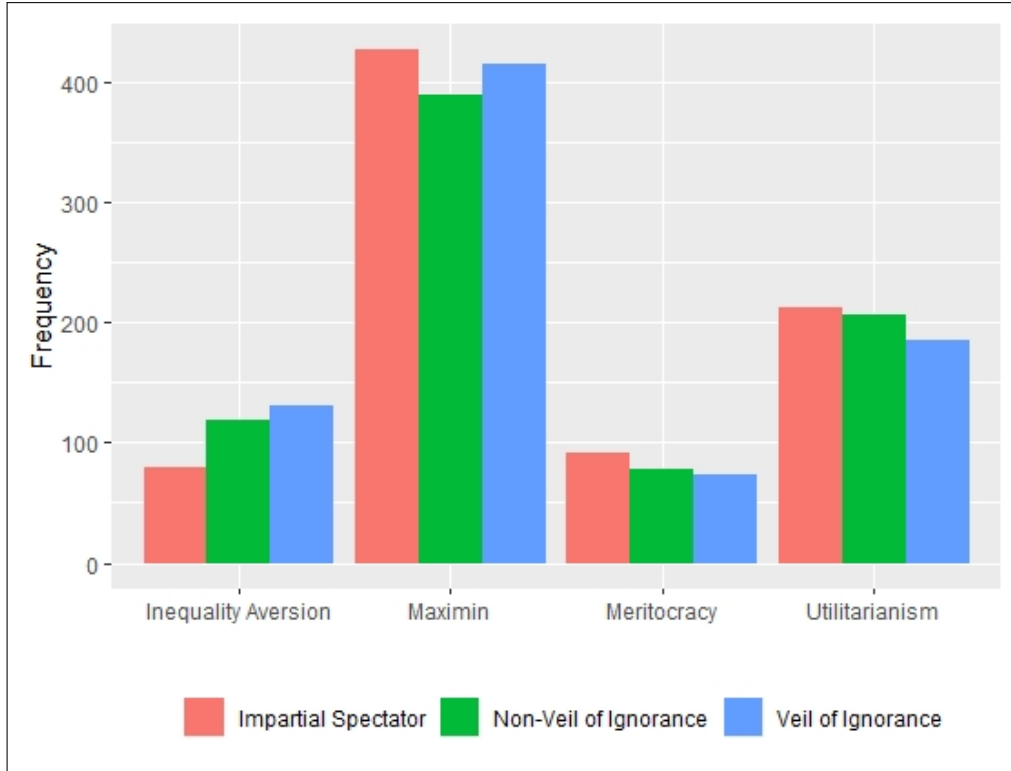
However, since we are concerned in H2 with whether norms help constitute individual social preferences (i.e. chosen principle), the more relevant norm is any perceived principle norm (i.e. Decision 2). In the aggregate, we now find that 55% of our subjects choose a principle that is the same as their perceived principle norm (i.e. the frequency is much greater than would be expected through chance if the decisions were independent of each other). Table 5 examines the individual level data by reporting on a regression where perceived principle norm is used as a predictor of principle choice. We find that knowing a person's perceived principle norm helps predict their social preference for all social norms. In online appendix section C we test the same model using the perceived distribution norm and also find that norms help predict individual social preferences, but with smaller coefficients than reported in table 5 and not for all social norms (the exception is utilitarianism).

Result 4 (supporting H2): there is evidence that principle social norms help predict social preferences at the individual and at the aggregate level.

We turn now to the elicitation mechanism hypotheses and whether the distribution decisions differ across the treatments. Figure 2 gives the aggregate frequency of distribution choices by treatment. There is a significant difference in chosen distributions by treatments (Chi-squared of 18.63, $p=0.05$); however, this is driven by Inequality Aversion as the significance disappears when we exclude this choice from the analysis (Chi-square of 2.89, $p=0.58$). Inequality Aversion is significantly less frequent under the Impartial Spectator than Veil and non-Veil procedures. This is what is also revealed by the treatment dummies in the individual level regression analysis of Table 4.

Result 5 (against H4): Distribution decisions are not skewed towards Maximin or Utilitarian/Efficient in Veil of Ignorance as compared with Impartial Spectator. The only skew is

Figure 2: Distribution choice by Treatment



towards Inequality Aversion in the Veil of Ignorance as compared with the Impartial Spectator.

Turning to H5 and the expected influence of selfishness on decisions, Table 4 reveals some influence from selfishness on decision making in the non-Veil treatment. However, it is not uniform in its effect in the sense of pushing decisions in the direction of selfishness. While the fact that a Meritocratic choice is in the person's selfish interest helps predict the choice of Meritocracy, the reverse is the case in Maximin choices: when Maximin is in the selfish interest, the person is less likely to select Maximin. Furthermore, the one significant aggregate difference in Figure 2 between the non-Veil treatment and the Impartial Spectator, where selfishness can play no role, is in the frequency of Inequality Aversion choices and yet the individual regressions in Table 4 do not suggest that selfishness influences the probability of selecting Inequality Aversion in the non-Veil treatment.

Table 5: Logistic regressions of social preferences for all treatments

	Inequality Aversion	Social preference		
		Maximin	Meritocracy	Utilitarianism
Social Norm	1.781*** (0.108)	2.160*** (0.152)	1.703*** (0.100)	1.849*** (0.116)
Constant	-1.652*** (0.415)	-2.346*** (0.553)	-0.460 (0.368)	-3.023*** (0.437)
Controls	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓
Session Fixed Effects	✓	✓	✓	✓
Observations	2,219	2,219	2,219	2,219
Pseudo R-squared	0.134	0.124	0.143	0.124

Notes: Estimates come from a logistic regression. Social Norm is a binary variable equal to 1 if the subject's perceived social norm in the principle choice matched the chosen principle. Robust standard errors are presented in parentheses. *** p<0.01, ** p<0.05 , * p<0.1.

Result 6 (against H5): Selfishness does not consistently help predict distribution decisions in non-Veil of Ignorance treatment.

Tables 6 gives the proportion of social preference followers and the proportion of perceived social norm followers by treatment. There are no significant differences in these frequencies across the three treatments (Chi-squared for preference following is 1.12, p=0.57 and for norm following 1.85, p=0.40). Furthermore, we know from the individual level regression in Table 4 that none of the treatment dummies are significant except for those who choose the inequality averse distribution.¹³

Result 7 (in support of H6): There are no significant differences in the frequency of social preferences-distribution consistency or norm-following-distribution consistency across the treatments.

¹³There is one further treatment difference that, although not germane to our hypotheses, is worth reporting. In Table 5 we find treatment effects in the likelihood that a particular principle will be chosen: ceteris paribus, the meritocratic principle is more likely to be selected in the impartial spectator treatment, while maximin and utilitarianism are more likely to be selected under the Veil of Ignorance.

Table 6: Preference- and Norm-following by Treatments

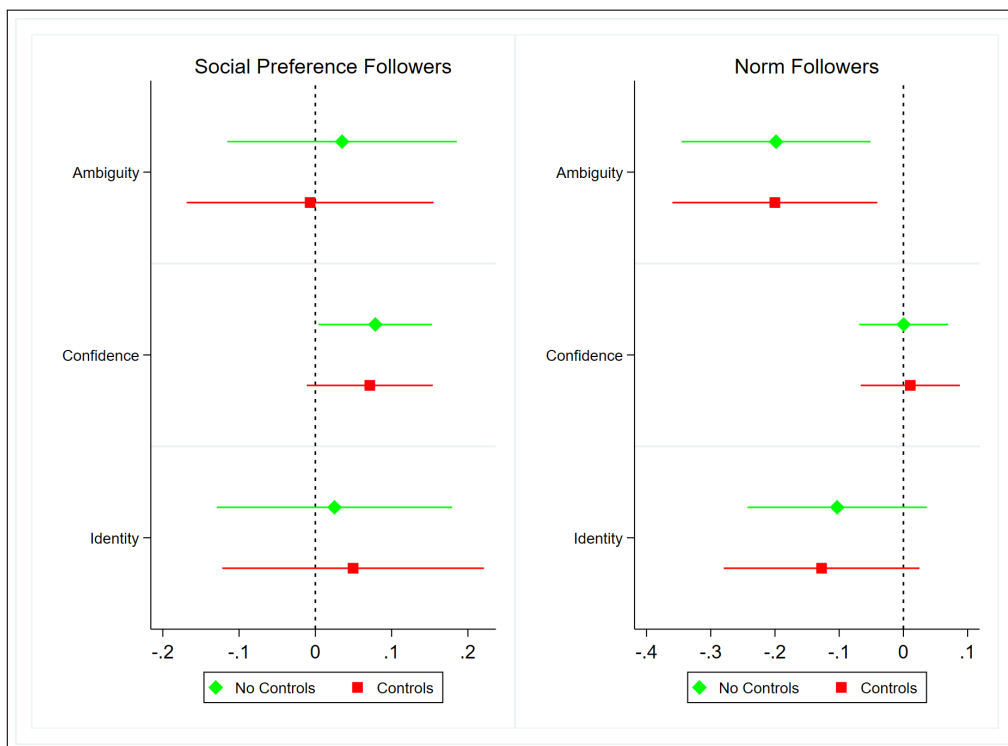
	Treatments		
	Impartial Spectator	Veil of Ignorance	Non-Veil of Ignorance
<i>Social Preference following</i>			
Inequality Aversion	18.78%	20.51%	17.62%
Maximin	76.24%	70.21%	60.19%
Meritocracy	19.21%	15.33%	14.53%
Utilitarianism	12.59%	13.33%	23.60%
<i>Social Norm following</i>			
Inequality Aversion	42.45%	50.66%	47.79%
Maximin	75.85%	77.90%	76.08%
Meritocracy	33.65%	43.84%	29.55%
Utilitarianism	52.27%	53.30%	54.75%

V ROBUSTNESS CHECKS

Since these results run strikingly counter to the conventional approach of explaining unselfish behavior by the presence of social preferences in a preference satisfying model of behavior, we decided to conduct a further robustness check experiment. We did this in two ways. Our immediate concern was to test for the possibility that, by asking subjects to identify their perceived distribution social norm immediately after they made the actual distribution decision, we might have rendered the distribution choice especially salient to the subjects when eliciting the distribution social norm. Thus, we inverted decision 3 and 4 in our robustness check experiment to produce the following order of decisions: subjects first identified their preferred principle and then their perceived principle norm. Decision 1 and 2 therefore remained in the same order as in the main experiment. We then asked subjects to identify their perceived distribution norm (decision 4) before making their distribution choice (decision 3). We find the same patterns as in our main results. In fact, they are a bit stronger in favour of norm-following (see online appendix B.1).

The second check on robustness came from exploring what distinguished norm-followers from

Figure 3: Individual Characteristics by Subject Group



Notes: Figures are based on logistic regressions. The outcome variable of the left coefficient plot is equal to 1 if the subject followed their social preference in the distribution choice and 0 otherwise. The outcome variable of the coefficient plot on the right is equal to 1 if the subject followed their perceived social norm and 0 otherwise. Ambiguity ranges from 0 to 7 (with a higher value indicating more ambiguity seeking preferences) and is a standardized scale based on the ambiguity preference survey module developed by [Cavatorta and Schröder \(2019\)](#). Confidence is measured as the subject's response to the question "On a scale from 1 to 10, please rate how confident you are in the choice you just made." which was asked directly after subjects chose a principle. A higher value indicates more confidence. Identity ranges from 1 to 4 with a higher value indicating a higher level of identity. This variable was measured using the module developed by [Kuo and Margalit \(2012\)](#).

those who acted according to their social preference or selfishly. In the robustness experiment, we asked subjects after they had chosen their distribution principle to assess on a Likert scale how confident they were in their choice. In so far, as they were not confident in this choice, we expect, on the basis of H3, that they would be more likely to follow a norm since lack of confidence plausibly reflects the kind of existential epistemic predicament that triggers norm-following. In the concluding demographic questions, we also included questions that were designed to elicit the subjects' social identification with groups, their ambiguity aversion and their tolerance of deception. Our conjecture was that if H3b explained norm-following

more than H3a, then ambiguity aversion would help predict norm-following. Likewise, a low tolerance of deception is sometimes argued to predict an inability to self-deceive and so would likely predict norm following if H3b explained this phenomenon (e.g. see [Trivers 2011](#)). Finally, we conjectured from social identification theory that those who identified with groups most strongly would be more likely to make confident choices of distribution principles.

As can be seen in the first plot of [Figure 3](#), we found that those who followed their social preference in the distribution choice expressed a higher level of confidence compared to all other subjects. In other words, a subject's level of confidence helps predict whether their distribution choices are consistent with a) their justice principle selection (when confidence is high), b) selfishness (when low) and c) norm-following (when low). Thus, confidence tends to split the population into social preference guided subjects on the one hand when confidence is high and either selfish or norm-followers on the other hand when confidence is low. This is consistent with H3a and H3b in the sense that confidence distinguished social preference guided subjects from those who follow norms. We also found that ambiguity aversion helps predict the likelihood of a subject being a norm-follower in their distribution choices. This can be seen in the second plot of [Figure 3](#) and suggests that H3b plays a significant role in explaining norm-following. Finally, we find that social identification helps predict confidence. Thus, there is some evidence that strong social identification helps explain why individuals act on social preferences (and of course, this is also consistent with [Result 4](#) where we find that principle norms, which might plausibly come from social identification, help predict individual social preferences). The full details of the second experiment and these further results can be found in the online appendix [A.3.5](#) and [C.4](#).

Both aspects of the robustness check provided by the second experiment, therefore, reinforce the conclusion that norm-following plays a more significant role in explaining unselfish

behavior than does the preference satisfying model with people acting on social preferences.

Our final set of robustness checks relates to the key assumption that we make with respect to individuals using principles of justice when thinking about how to make distribution decisions. In particular, this is crucial in making the connection between individual’s chosen principle of justice and their likely social preferences. At the end of the experiment we asked our subjects in an open commentary box to explain how they decided on their distribution option. Table 7 lists the most frequently used terms by chosen distribution.

Table 7: Terms most Frequently used to Justify chosen Distribution

Inequality Aversion				Maximin			
	Total Frequency	Documents	Relative		Total Frequency	Documents	Relative
Equal distribution	5	5	0.011	Hard work	13	13	0.011
Income inequality	4	4	0.009	Equal distribution	12	12	0.010
Basic income	3	3	0.007	Fair distribution	10	10	0.008
Equal amount	3	2	0.004	Greater good	8	8	0.006
Distribute wealth	3	3	0.007	Income inequality	8	8	0.006
Shared equally	2	2	0.004	Many people	7	5	0.004
Best choice	2	2	0.004	make sure	7	7	0.006
Fair distribution	2	2	0.004	Income distribution	7	7	0.006
Observations	447	447	447		1,231	1,231	1,231
Meritocracy				Utilitarianism			
	Total Frequency	Documents	Relative		Total Frequency	Documents	Relative
Work hard	5	5	0.018	Work hard	12	9	0.015
Felt right	3	3	0.011	Work harder	6	5	0.008
Work harder	3	3	0.011	Hard work	6	6	0.010
worked harder	2	2	0.007	Paid based	4	4	0.007
Hard work	2	2	0.007	Seemed fair	4	4	0.007
Next group	2	1	0.004	Felt right	3	3	0.005
Make sure	2	2	0.007	Worked hard	3	3	0.005
Second game	1	1	0.004	Worth taking	3	3	0.005
Observations	271	271	271		603	603	603

Notes: The table reports the most frequently used terms used by respondents to justify their chosen distribution. Total frequency reports the number of times a term was used overall within the subgroup of respondents who chose a particular distribution. Documents reports the number of responses of individual respondents in which a term was used at least once. Relative reports the proportion of responses within the distribution-dependent subgroup that refer to the given term.

The most used words differ substantially for each distribution choice and, importantly, match the wording of our principle options. This is particularly striking when comparing the terms used to justify the inequality averse and maximin distributions with the meritocratic and utilitarian distributions. In short, the currency that people use to explain their decisions is the same as that of the principles of justice, even though, as we have seen they are not

typically guided by such principles.

Another possible qualification to our conclusion might be that our subjects are guided by more than one justice principle and it is possible that a different secondary principle of justice was triggered when the actual distribution choices were presented in Decision 3. We ran a further survey of 200 subjects where we asked them after they had identified the principle of justice they thought should be applied (Decision 1) if they had a secondary justice principle, and if so, what it was. Just over half (56%) had a secondary principle and of those who did, maximin was again the least chosen (secondary) principle (see appendix C.5.1). Less than 9% of the 200 subjects identified Maximin as their secondary principle and so the possible contribution of a secondary principle in explaining the wholesale shift to Maximin in the distribution Decision 3 is at best relatively modest even if all these 9% had been guided by their secondary principle alone. Recall in the original experiment 14% identified Maximin as their principle and 50% chose the Maximin distribution.: even another 9% leaves a big gap.

The final possible qualification that we considered was that, although each principle does identify one of the four distribution outcomes, subjects might have made an execution error when translating their principle into an actual distribution decision. Random ‘trembling’ would, however, introduce ‘noise’ and weaken the principle-distribution consistency (as it might any norm-distribution consistency); it would not explain why the distribution decisions are actually skewed to the Maximin distribution. For this to occur there has to be some reason for supposing that ‘errors’ are easier to make in the Maximin direction because Maximin is ‘closer’ to each of the principles than is any of the others. We test for this possibility by asking another survey of 200 subjects to choose a principle (i.e. Decision 1) and then we ask them to identify the distribution (in Decision 3) that they associate with their chosen principle. Those who incorrectly identify their chosen principle’s distribution do on average err noticeably in the direction of two distribution outcomes: 44% go to Utilitarianism and

41% go to Maximin. Most (82%) of the trembles to Maximin were accounted for by those who identified their chosen principle as Inequality Aversion, so we re-ran the individual regression in Table 4 excluding all the subjects who chose the Inequality Aversion principle in Decision 1. The perceived social norm is still a more important predictor of these remaining subjects' distribution choices than is their chosen principle (see appendix C.6.2). So, while 'skewed' trembling might explain why those who chose Inequality Aversion migrated to the Maximin distribution, it does not explain why this occurs for subjects that select the other principles (and they are the majority in our sample). Indeed, the errors among the subjects choosing Meritocracy (our modal principle choice) were skewed away from Maximin (only 8% of their mistakes went to Maximin).¹⁴

In short, after a variety of robustness checks, our key result still holds: on balance our subjects shift to the Maximin distribution outcome most likely because they typically identify and are guided by Maximin as the social norm.

VI DISCUSSION AND CONCLUSION

There are several respects in which the behavior of our subjects is reassuringly consistent with other experimental findings. For example, we find in Table 4 that being trained in economics is a powerful predictor of choosing the Utilitarian/Efficient distribution but not any of the other distributions; and we know, for example, from [Fehr et al. \(2006\)](#) that economics students are more inclined to be influenced by efficiency considerations than non-economics students in such distribution decisions. Likewise, it is known that US subjects hold more meritocratic beliefs than European subjects (see [Alesina and Glaeser 2004](#)) and we too find in Table 4 that the only predictable difference from nationality is that being a US citizen

¹⁴It is perhaps also worth noting that the trembling rate was over 50%: that is only 45% correctly identified the distribution outcome associated with their chosen principle. Again, this suggests that the majority of our subjects were not used to thinking in terms of principles of justice; and if this is the case, it would be difficult for the majority of our subjects to be said to have social preferences that they consult when decision making in this instance.

increases the probability of selecting the Meritocratic distribution. Being to the right on a typical left-right political question regarding the role of government in the economy helps predict the Utilitarian/efficient distribution; whereas being on the left helps predict Maximin. This is in line with the common finding that a left-leaning political orientation is associated with a preference for more redistribution (see [Alesina and Giuliano 2011](#)). Again, being risk averse helps predict Maximin, as would be expected. Finally, our evidence on the influence of social norms is consistent with what has been found in other studies (e.g. [Krupka and Weber 2013](#); [Kimbrough and Vostroknutov 2016](#)).

Our key contribution is to put preference satisfaction in competition with the norm-following one as a psychological account of unselfish behavior. We find evidence for both in the individual level data (Result 1 and 2). But social norms have a quantitatively bigger effect than social preferences in this individual level data and while social norms are useful predictors in the aggregate data, there is no evidence in the aggregate data for the influence of social preferences. So, although, in practice, it is not ‘either/or’, the evidence is stronger for norm-following (Results 3): indeed, knowing whether someone is a norm-follower is more than twice as important as knowing their preferred principle in predicting their distribution decision in our experiment. Furthermore, there is some evidence that norms may help constitute social preferences (Result 4) and so some part of the influence that we associate with social preferences may ultimately also follow from that of social norms.

On the choice of elicitation mechanism, we find that it makes surprisingly little difference to the observed distribution choices. For example, the Veil of Ignorance does not, as would be expected, produce a significantly higher fraction of Maximin choices than the Impartial Spectator. Nor does selfishness appear consistently to help explain choices in the non-Veil of Ignorance treatment. Finally, there is no evidence that the choice of elicitation mechanism affects our first question: the relative consistency between distribution choices and social

preferences on the one hand and social norms on the other.

These conclusions with respect to the elicitation mechanism are broadly consistent with the first finding on the relative importance of social norms as compared with social preferences. This is because the hypotheses regarding the elicitation mechanisms are built around the social preferences model and the way that each potentially combines selfish and social preferences in different ways. However, if distribution decisions are generally better explained by norm-following behavior, then there is not the same reason for supposing that the elicitation procedures will differ systematically in the manner suggested by H4 and H5. Indeed, H6 and Result 7 suggest that the elicitation procedure does not affect the finding that norms contribute more to distribution decisions than social preferences.

These results are important in four respects.

First, they suggest that the use of the Pareto principle in welfare economics has a weak foundation whenever people behave unselfishly because such unselfish behavior is not well captured by a preference satisfying model in our experiment. In particular, it cannot be assumed that unselfish behavior reveals social preferences which can then be entered into a social welfare function for the purposes of developing policy recommendations.

Second, this, in turn, means that the foundations of welfare economics need reworking to take account of norm-following. This is non-trivial because we have an experiment where the influence of social preferences is carefully distinguished from that of social norms. The support for H3 on norms is in favour of a kind of norm-following which cannot be reduced or re-described as a kind of social preference guided behavior. This need not be deeply antithetical to the preference satisfying model because there are, for example, evolutionary explanations of norms that cast them as shared behaviors that enhance individual fitness. Nevertheless, it poses a significant challenge for welfare economics.

Third, we have some insights into why people might be guided by norms rather than social preferences. There is evidence that it arises from an epistemic problem with respect to what preferences to act upon. Those who lack confidence in their chosen principle and who are ambiguity averse are inclined to follow their perceived norm. Interestingly, those who have high confidence and so are more likely to be guided by their social preferences also tend to have high levels of social identification.

Finally, it may be possible to draw some useful substantive insights with respect to the character of unselfish behavior from this experiment. Some care is required because we only have four actual distributions and had the option set been different, then there might have been different choices. Furthermore, the character of the unselfish behavior that is revealed may depend on the initial distribution of income that we have assumed. Nevertheless, the average EU actual top 20%/bottom 20% ratio for disposable income is very close to the 5.5 we have assumed (see [Eurostat 2018](#)). So, in this respect, the decision problem captures something close to the current post tax relativities and may be relevant to the contemporary discussion regarding how further intervention might be required to alter the income distribution. For example, both the IMF ([Ostry et al. 2014](#)) and OECD ([OECD 2015](#)) have argued that a move to greater equality would in current circumstances help to boost productivity growth. In this context, our experiment suggests that the majority reveal support for policies that improved the position of the bottom 20%.

REFERENCES

- AKERLOF, G. A. AND R. E. KRANTON (2000): “Economics and identity,” *The quarterly journal of economics*, 115, 715–753.
- ALESINA, A. AND P. GIULIANO (2011): “Preferences for redistribution,” in *Handbook of social economics*, Elsevier, vol. 1, 93–131.
- ALESINA, A. AND E. L. GLAESER (2004): *Fighting poverty in the US and Europe: A world of difference*, Oxford University Press.
- ALGER, I. AND J. W. WEIBULL (2013): “Homo moralis—preference evolution under incomplete information and assortative matching,” *Econometrica*, 81, 2269–2302.
- ANDREONI, J. AND J. MILLER (2002): “Giving according to GARP: An experimental test of the consistency of preferences for altruism,” *Econometrica*, 70, 737–753.
- APESTEGUIA, J., S. HUCK, AND J. OECHSSLER (2007): “Imitation—theory and experimental evidence,” *Journal of Economic Theory*, 136, 217–235.
- BERNHEIM, B. D. (2009): “Behavioral welfare economics,” *Journal of the European Economic Association*, 7, 267–319.
- BICCHIERI, C. (2005): *The grammar of society: The nature and dynamics of social norms*, Cambridge University Press.
- BINMORE, K. (2010): “Social norms or social preferences?” *Mind & Society*, 9, 139–157.
- BOLTON, G. E. AND A. OCKENFELS (2000): “ERC: A theory of equity, reciprocity, and competition,” *American economic review*, 90, 166–193.
- (2006): “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment,” *American Economic Review*, 96, 1906–1911.
- BREGMAN, R. (2020): *Humankind: A hopeful history*, London: Bloomsbury.
- CAPPELEN, A. W., J. KONOW, E. Ø. SØRENSEN, AND B. TUNGODDEN (2013): “Just luck: An experimental study of risk-taking and fairness,” *American Economic Review*, 103, 1398–1413.
- CAVATORTA, E. AND D. SCHRÖDER (2019): “Measuring ambiguity preferences: A new ambiguity preference survey module,” *Journal of Risk and Uncertainty*, 58, 71–100.
- CHARNESS, G. AND M. RABIN (2002): “Understanding social preferences with simple tests,” *The Quarterly Journal of Economics*, 117, 817–869.
- CHAUDHURI, A. (2011): “Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature,” *Experimental economics*, 14, 47–83.
- DELLAVIGNA, S., J. A. LIST, U. MALMENDIER, AND G. RAO (2020): “Estimating social preferences and gift exchange with a piece-rate design,” *CEPR Discussion Paper No. DP14931*.

- DUESENBERY, J. ET AL. (1960): “Comment on “An economic analysis of fertility,”” *Demographic and economic change in developed countries*, 231–34.
- DURANTE, R., L. PUTTERMAN, AND J. VAN DER WEELE (2014): “Preferences for redistribution and perception of fairness: An experimental study,” *Journal of the European Economic Association*, 12, 1059–1086.
- DURKHEIM, É. (2013): *Durkheim: The division of labour in society*, Macmillan International Higher Education.
- ELLINGSEN, T., M. JOHANNESSON, J. MOLLERSTROM, AND S. MUNKHAMMAR (2012): “Social framing effects: Preferences or beliefs?” *Games and Economic Behavior*, 76, 117–130.
- ENGELMANN, D. AND M. STROBEL (2004): “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments,” *American economic review*, 94, 857–869.
- ENKE, B. (2019): “Kinship, cooperation, and the evolution of moral systems,” *The Quarterly Journal of Economics*, 134, 953–1019.
- EUROSTAT (2018): “Income Inequality in the EU,” Accessed: 20/03/2020, <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180426-1>.
- EYAL, P., R. DAVID, G. ANDREW, E. ZAK, AND D. EKATERINA (2021): “Data quality of platforms and panels for online behavioral research,” *Behavior Research Methods*, 1–20.
- FATAS, E., S. P. H. HEAP, AND D. R. ARJONA (2018): “Preference conformism: An experiment,” *European Economic Review*, 105, 71–82.
- FEHR, E., M. NAEF, AND K. M. SCHMIDT (2006): “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment,” *American Economic Review*, 96, 1912–1917.
- FEHR, E. AND K. M. SCHMIDT (1999): “A theory of fairness, competition, and cooperation,” *The quarterly journal of economics*, 114, 817–868.
- FISMAN, R., I. KUZIEMKO, AND S. VANNUTELLI (2020): “Distributional preferences in larger groups: Keeping up with the Joneses and keeping track of the tails,” *Journal of the European Economic Association* (forthcoming).
- GÄCHTER, S., D. NOSENZO, AND M. SEFTON (2013): “Peer effects in pro-social behavior: Social norms or social preferences?” *Journal of the European Economic Association*, 11, 548–573.
- GINTIS, H. (2010): “Social norms as choreography,” *politics, philosophy & economics*, 9, 251–264.
- GUALA, F., L. MITTONE, AND M. PLONER (2013): “Group membership, team preferences, and expectations,” *Journal of Economic Behavior & Organization*, 86, 183–190.
- HARSANYI, J. C. (1955): “Cardinal welfare, individualistic ethics, and interpersonal com-

- parisons of utility,” *Journal of political economy*, 63, 309–321.
- (1980): “Can the maximin principle serve as a basis for morality? A critique of John Rawls’s theory,” in *Essays on ethics, social behavior, and scientific explanation*, Springer, 37–63.
- HOLLIS, M. ET AL. (1994): *The philosophy of social science: An introduction*, Cambridge University Press.
- KIMBROUGH, E. O. AND A. VOSTROKNUTOV (2016): “Norms make preferences social,” *Journal of the European Economic Association*, 14, 608–638.
- KRUPKA, E. L. AND R. A. WEBER (2013): “Identifying social norms using coordination games: Why does dictator game sharing vary?” *Journal of the European Economic Association*, 11, 495–524.
- KUO, A. AND Y. MARGALIT (2012): “Measuring individual identity: Experimental evidence,” *Comparative Politics*, 44, 459–479.
- LEDYARD, O. (1995): “Public goods: some experimental results,” *Handbook of experimental economics*, 1.
- NOZICK, R. (1974): *Anarchy, state, and utopia*, vol. 5038, New York: Basic Books.
- OECD (2015): “Growth and income inequality: trends and policy implications,” *OECD Economics Department Policy Notes*.
- OSTRY, M. J. D., M. A. BERG, AND M. C. G. TSANGARIDES (2014): *Redistribution, inequality, and growth*, International Monetary Fund.
- PARSONS, T. ET AL. (1949): *The structure of social action*, vol. 491, Free press New York.
- PATERNOTTE, C. AND J. GROSE (2013): “Social norms and game theory: Harmony or discord?” *The British journal for the philosophy of science*, 64, 551–587.
- PEER, E., L. BRANDIMARTE, S. SAMAT, AND A. ACQUISTI (2017): “Beyond the Turk: Alternative platforms for crowdsourcing behavioral research,” *Journal of Experimental Social Psychology*, 70, 153–163.
- RAWLS, J. (1971): *A theory of justice*, Harvard university press.
- SMITH, A. (1759): “The theory of moral sentiments, ed,” *A. Millar, A. Kincaid & J. Bell. AM Kelley. (Originally published in 1759).*
- TAJFEL, H., J. C. TURNER, W. G. AUSTIN, AND S. WORCHEL (1979): “An integrative theory of intergroup conflict,” *Organizational identity: A reader*, 56, 65.
- TRIVERS, R. (2011): *The folly of fools: The logic of deceit and self-deception in human life*, Basic Books (AZ).

A MATERIALS AND METHODS

A.1 Overview

We conducted our online experiment using Qualtrics for the design of the study and Prolific Academic for the recruitment of participants. Prolific Academic is a web-based panel with about 300,000 participants as of October 2021. Participants on Prolific have been found to pay significantly more attention and provide responses of higher quality than those registered on mTurk (Peer et al. 2017; Eyal et al. 2021).

Our main experiment was conducted on the 14th of November and the 9th of December 2019. The average completion time was 8 minutes and 17 seconds and respondents earned on average £1.55 for their participation. The full survey instrument that we used is available in Section E of this appendix. The data and code used for the analysis will be made available online at Harvard’s Dataverse for replication purposes upon acceptance for publication.

A.2 Sampling and Survey Implementation

We conducted a total of two main waves of the experiment, as well as seven additional waves for robustness checks. Table 1 provides an overview of all waves.

We focused our online experiment on participants from the US, UK and the following Western European countries: Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, Netherlands, Norway, Sweden and Spain. Table 2 lists the number of respondents from each geographical area by individual wave. To ensure that we reached respondents from all geographical areas, all waves were ran in the late afternoon GMT time. Our samples are not representative of individual countries. Descriptive statistics of the sample composition can be found in section B.

A.3 Survey Structure

A.3.1 Basic Set up

Introduction

Subjects are asked for their consent to participate in the study and reminded to read the questions very carefully and answer honestly.

Table 1: Overview of individual waves

	Date	Sample Size	Avg. Time	Returned	Timed Out
First Wave	14/11/2019	1,205	8.11mins	27	16
Second Wave	09/12/2019	1,203	7.45mins	32	25
Average Income Test	30/03/2020	294	14.05mins	59	15
Social Norm Test	30/03/2020	302	11.00mins	36	3
Motivation Test	21/04/2020	1,003	15.08mins	67	37
Second Principle Test	19/11/2020	201	3.48mins	4	2
Distribution Test	25/11/2020	200	4.37mins	5	1
Wording Test	08/10/2021	222	4.14mins	12	1
Order Test	08/10/2021	218	9.52mins	21	1

Table 2: Sample composition of individual waves

	United Kingdom	United States	Western Europe	Total Sample Size
First Wave	768	165	272	1,205
Second Wave	623	280	300	1,203
Average Income Test	153	65	76	294
Social Norm Test	180	48	72	302
Motivation Test	561	48	392	1,003
Second Principle Test	18	120	63	201
Distribution Test	84	9	107	200
Wording Test	89	38	95	222
Order Test	76	42	100	218

Experimental Part

Using Qualtrics' *Randomizer*, subjects are randomly and evenly allocated to one of three treatments for the following four decisions.

Decision 1. Identify guiding principle of justice.

Decision 2. Incentivised guess of what decision most people made in Decision 1.

Decision 3. Select distribution.

Decision 4. Incentivised guess of what decision most people made in Decision 3.

Quiz

Demographic Questions

A.3.2 Treatments

Different institutional mechanisms for eliciting justice principles and making distribution decisions (each encoding a different idea over how best to identify what is just).

Treatment 1: *Impartial Spectator*. Decision 1-4 undertaken as an impartial spectator.

Treatment 2: *Veil of Ignorance*. Decision 1-4 undertaken behind a veil of ignorance.

Treatment 3: *Non-veil of Ignorance*. Decision 1-4 undertaken knowing one's own likely position in the distribution.

A.3.3 Robustness Check 1: Average Income Test

In the main two waves of the experiment we referred to "Total" income per distribution choice. We therefore conducted a robustness check where we replaced "Total" with "Average" in all displays of our distribution options.

A.3.4 Robustness Check 2: Social Norm Test

The [Krupka and Weber \(2013\)](#) method uses a separate subject pool to elicit the social norm for a particular decision problem. Our main experiment uses the same subject pool for norm elicitation and so we conducted an additional norm elicitation experiment with a separate subject pool. This experiment only consisted of decision 4 of the experimental part outlined in section A.3.1.

A.3.5 Robustness Check 3: Motivation Test

Our main robustness check was designed to test the motivations behind norm following and included the following elements in addition to the main experiment:

- **Ambiguity preference elicitation.** We followed the method developed by [Cavatorta and Schröder \(2019\)](#) to measure subjects' ambiguity preferences.
- **Confidence in principle.** After subjects made decision 1, they were asked to rate their confidence in the chosen principle: *On a scale from 1 to 10, please rate how confident you are in the choice you just made.*

- **Identity elicitation.** Following [Kuo and Margalit \(2012\)](#) we asked respondents the following two additional questions in the demographics section:

1. Some people describe themselves by their nationality, their ethnicity, their race, their religion, or their occupation. How about you? Do you identify first and foremost by:
 - Your nationality
 - Your ethnicity
 - Your race
 - Your religion
 - Your occupation
 - Other (Please specify)
2. Consider your response to the previous question. How strong would you say your attachment is to the identity you chose? Would you say your attachment is:
 - Not strong at all
 - Slightly strong
 - Somewhat strong
 - Very strong

- **Self-deception elicitation.** To elicit subjects' level of self-deception we asked the following two additional questions in the demographics section:

1. It has been argued that there will always be occasions when the kindest thing to do is lie. But, on the other hand, if people lie, then who can you believe? Do you agree it is okay to lie sometimes?
 - Scale ranges from 1 (Strongly agree) to 7 (Strongly disagree)
2. There is a big debate in psychology over whether deception in experiments should be permitted. What do you think?
 - Scale ranges from 1 (Never) to 7 (Whenever it helps science)

We further reversed the order of decision 3 and 4 in this robustness check to test whether people simply chose the same distribution option in decision 4 that they chose in decision 3, for example, to appear consistent. The results in section C.4.3 confirm that this was not the case. This robustness check also only included the impartial spectator treatment as we did not find significant treatment effects in our main waves.

A.3.6 Robustness Check 4: Second Principle Test

To test for the possibility that our subjects have two principles that they take into consideration when making the distribution choice we conducted a further robustness check asking subjects first, whether they had another principle they agreed with and second, which of the other principles it is.

A.3.7 Robustness Check 5: Distribution Test

To ensure that subjects understood which distribution option corresponded to which justice principle we conducted a robustness check asking subjects to identify the distribution corresponding to their chosen principle. This decision was incentivised. If subjects correctly identified the corresponding distribution they received a bonus payment of 50p.

A.3.8 Robustness Check 6: Wording Test

As pointed out by one referee, the wording of our principle statements is not structured in an entirely consistent manner which could have affected subjects' likelihood to choose one principle over another. To test for this possibility, we conducted a robustness check with an alternative wording of the inequality aversion and maximin statements. We also repeated the distribution test introduced in robustness check 5 to check whether subjects are more or less likely to correctly identify the distribution corresponding to their chosen principle given this alternative wording. The wording used in this test is as follows:

Maximin: Income should be distributed to improve the position of the least well-off group in society.

Inequality Aversion: Income should be distributed to reduce inequality by minimizing average differences in income.

A.3.9 Robustness Check 7: Order Test

While we already reversed the order of decisions 3 and 4 in robustness check 2, we added a seventh robustness check to reverse the order of decisions 1 & 2 and 3 & 4. This allows us to test whether making the distribution decision first affects either the chosen distribution and principle, preference consistency, or norm-following.

B ADDITIONAL DESCRIPTIVE RESULTS

Table 3 reports summary statistics of all waves of the study. Our sample is clearly skewed towards younger respondents on low income. Over 50% of our sample has an annual income below £20,000. Except for the Average Income Test, our sample is also predominantly female.

Table 3: Summary Statistics of Demographics by Wave

	Main Experiment	Average Income Test	Social Norm Test	Motivation Test	Second Principle Test	Distribution Test	Wording Test	Order Test
<i>Demographics (%)</i>								
Female	60.10	49.32	56.61	60.10	47.96	52.53	55.07	48.10
Age								
18-20	9.58	13.65	14.67	14.34	15.58	18.09	10.96	14.49
21-29	35.47	41.98	36.00	43.03	48.24	47.74	40.64	42.99
30-39	28.85	24.91	24.33	25.68	21.11	19.10	28.31	26.64
40-49	13.61	12.63	13.00	10.63	8.54	10.55	14.16	11.21
50-59	8.45	4.78	9.00	5.12	5.53	2.01	3.65	2.34
60+	4.04	2.05	3.00	1.20	1.01	2.51	2.28	2.34
Students	24.92	27.55	29.33	31.70	38.31	34.50	33.78	35.94
Economics	21.47	29.33	21.67	21.38	27.00	26.00	21.62	27.19
Income								
Under £20,000	51.69	50.36	51.60	53.76	58.15	46.84	36.63	38.05
£20,000 to £34,999	25.74	23.36	25.98	27.21	23.37	30.38	33.17	31.22
£35,000 to £44,999	11.69	11.68	10.32	12.17	11.96	17.72	13.86	15.12
£50,000 to £74,999	6.65	7.66	7.12	4.87	3.80	2.53	9.41	9.76
£75,000 to £99,999	2.05	2.19	2.85	1.00	1.09	2.53	4.46	3.90
Over £100,000	2.19	4.74	2.14	1.00	1.63	0.00	2.48	1.95
Sample								
United Kingdom	57.77	52.04	60.00	56.04	8.96	42.00	40.09	34.86
United States	18.48	22.11	16.00	4.80	59.70	4.50	17.12	19.27
Europe	23.75	25.85	24.00	39.16	31.34	53.50	42.79	45.87
Observations	2,408	294	302	1,003	201	200	222	218

Table 4 reports descriptive variables by assigned treatment for our main experiment consisting of the first and second wave of the experiment. Most demographics are well-balanced between the treatment groups; however, the proportion of economics students is significantly different across treatment groups. Given that this variable does not appear to influence choices in the main variables of interest, this does not appear to be a problem for inference. The table further reports that quiz performance is significantly higher in the Non-Veil of Ignorance treatment. This is likely to be the case as respondents in this treatment answered two sample quiz questions prior to making their distributive decisions and were therefore better prepared for the actual quiz than respondents in the other two treatments. This significant difference however equally does not affect our main variables of interest.

B.1 Distribution of Main Variables

Figures 1 and 2 report the distribution of respondents' stated social preference, social norm and chosen distribution for the average income and motivation test, respectively. Both distri-

Table 4: Balance across treatment groups

	Impartial Spectator	Non-Veil of Ignorance	Veil of Ignorance
<i>Mean values</i>			
Female	59.60	61.02	59.70
Age	8.51	9.90	10.34
18-20	38.10	34.14	34.12
21-29	38.10	34.14	34.12
30-39	26.76	30.58	29.27
40-49	13.32	14.47	13.08
50-59	9.00	8.12	8.22
60+	4.32	2.79	4.98
Students	24.54	23.61	26.58
Economics	18.74**	24.12**	21.61
Income			
Under £20,000	49.80	51.15	54.13
£20,000 to £34,999	26.76	23.89	26.53
£35,000 to £44,999	13.32	12.15	9.60
£50,000 to £74,999	5.59	8.64	5.73
£75,000 to £99,999	2.40	2.16	1.60
Over £100,000	2.13	2.02	2.40
Sample			
United Kingdom	58.32	56.69	58.26
United States	17.02	19.32	19.13
Europe	24.66	23.99	22.61
Left-Right	4.03	4.03	4.02
Risk preference	5.58	5.60	5.42
Quiz performance	2.25	2.55***	2.31
Observations	811	792	805

Notes: Table reports the mean values for each variable. Asterisks indicate significant differences in mean values between treatment groups from a chi-squared test of independence. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

butions show a strikingly similar pattern. Meritocracy is the most chosen social preference, yet Maximin is by far the most chosen distribution and perceived social norm. In both distributions it is also evident that distribution choices are more closely aligned with perceived social norms than social preferences.

Figures 3 and 4 report the distribution of respondents' stated social preference, social norm and chosen distribution for the wording and order tests, respectively. Here, Maximin is again the most chosen distribution and Meritocracy the modal social preference in both tests. While Maximin is also the most chosen perceived social norm in the wording test, this is not the case in the order test. Here, utilitarianism is, in fact, the modal perceived social norm. Importantly however, the difference between the number of respondents who chose Maximin and those who chose Utilitarianism as their perceived social norm is only seven out of 218, suggesting that this finding, which is inconsistent compared to all other robustness checks, might be due to sampling.

Figure 1: Distribution of Social Preference, Norm and Distribution in Average Income Test

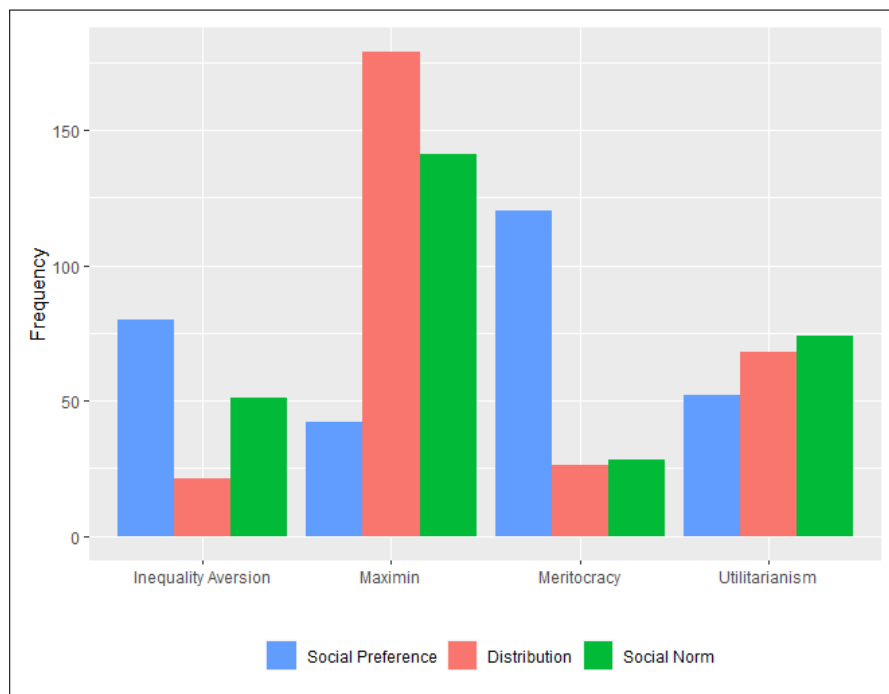
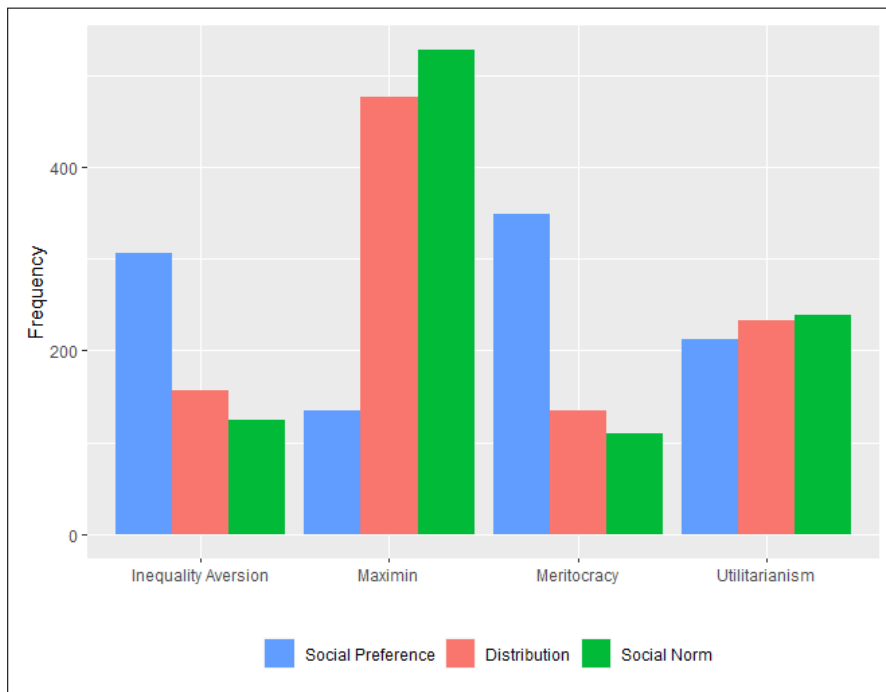


Figure 2: Distribution of Social Preference, Norm and Distribution in Average Income Test



C ADDITIONAL RESULTS

C.1 Main Experiment

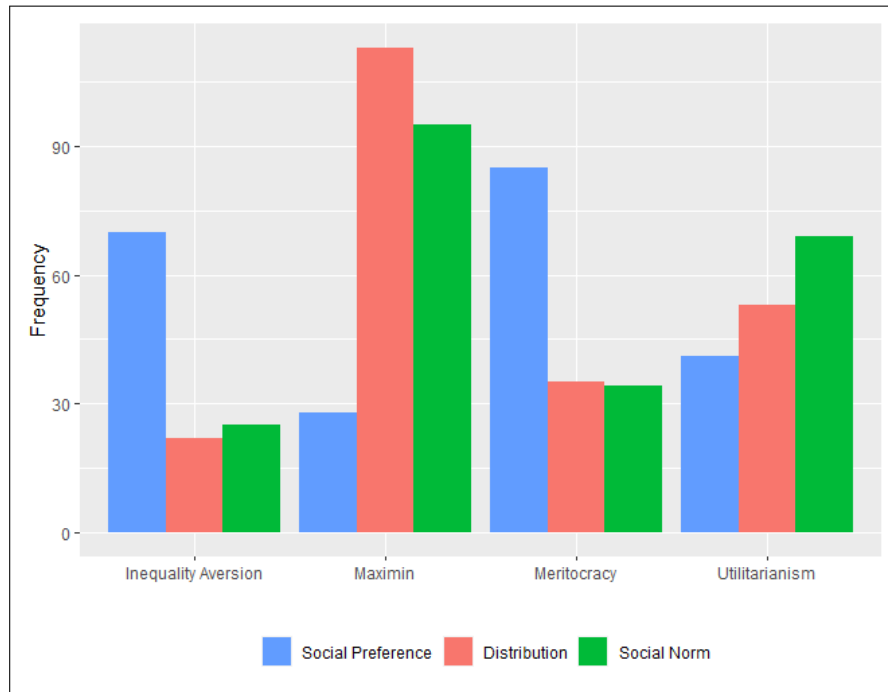
Table 5 reports logistic regressions similar to table 5 in the main text, whereby perceived social norm is used as a predictor of principle choice. Here, the perceived social norm is related to the distribution and not the principle choice. We find that knowing a person’s perceived distribution norm helps predict their social preference for three out of four social norms, the exception being Utilitarianism. The coefficients are also significantly smaller than for the models that use the principle norm as the explanatory variable.

C.2 Average Income Test

C.2.1 Preference- and Norm-following by chosen Distribution

Table 6 reports the chosen distribution by stated social preference and social norm for respondents in the Average Income Test. The pattern visible in table 6 is similar to the results of the main experiment: Social Norms are more closely related to distribution choices than social preferences, except for respondents who chose the Maximin distribution.

Figure 3: Distribution of Social Preference, Norm and Distribution in Wording Test



C.2.2 Main results

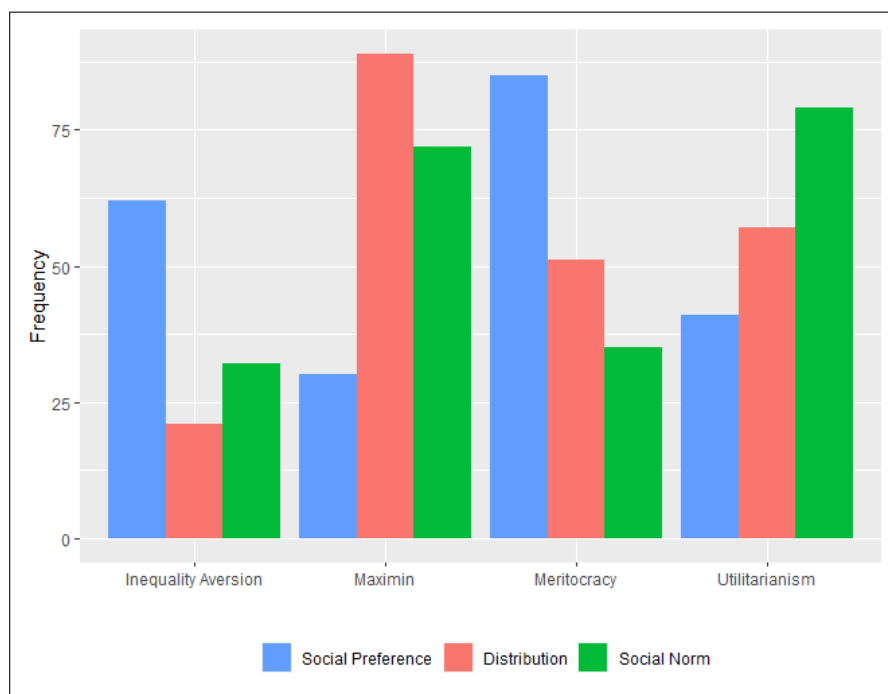
Table 7 reports the results of logistic regressions with individual distribution choices as the outcome variables for respondents in the Average Income Test. This test was conducted with only the Impartial Spectator treatment. These regression results are directly comparable to table 4 in the main text. Despite the small sample size of this robustness check, social norms are a highly significant predictor of distribution choices while social preferences only matter for the distribution choices of respondents who chose the meritocratic distribution. The social norm coefficients are similar to those of the main regression results.

C.3 Social Norm Test

C.3.1 Distribution of perceived Social Norm

Figure 5 reports the frequency of the perceived social norm of subjects in the Social Norm Test. Unlike in all our other waves, Utilitarianism is the modal choice while Maximin is the second most-frequent choice. As this distribution is strikingly different to all other waves of the experiment, it suggests that the respondents make a substantially different choice when asked to decide on the appropriate social norm for a separate group of subjects (as proposed by [Krupka and Weber 2013](#)) than when the decision is made on the same subject group.

Figure 4: Distribution of Social Preference, Norm and Distribution in Order Test



C.4 Motivation Test

C.4.1 Motivation by Subject Group

Table 8 reports individual characteristics for respondents who followed their social preference and those who followed their perceived norm in the distribution choice. While confidence in the chosen principle increases preference-following, more ambiguity aversion (a lower ambiguity preference score) is associated with norm-following. Interestingly, identifying with one's own race significantly decreases the likelihood of following one's perceived social norm. Table 9 reports individual predictors of respondents' confidence in their chosen principle. A stronger social identity is thereby associated with a higher level of confidence in one's chosen principle.

C.4.2 Preference- and Norm-following by chosen Distribution

Table 10 reports the chosen distribution by stated social preference and social norm for respondents in the Motivation Test. The pattern visible in table 10 is again similar to the results of the main experiment: Social Norms are more closely related to distribution choices than social preferences, except for respondents who chose the Maximin distribution. The proportion of respondents who chose the distribution that matches their perceived social

Table 5: Logistic regressions of social preferences for all treatments

	Inequality Aversion	Social preference		
		Maximin	Meritocracy	Utilitarianism
Social Norm	0.363*** (0.128)	0.431*** (0.122)	0.635*** (0.141)	-0.063 (0.119)
Constant	-1.176*** (0.386)	-2.210*** (0.523)	0.236 (0.339)	-2.312*** (0.396)
Controls	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓
Session Fixed Effects	✓	✓	✓	✓
Observations	2,219	2,219	2,219	2,219
Pseudo R-squared	0.024	0.026	0.043	0.012

Notes: Estimates come from a logistic regression. Social Norm is a binary variable equal to 1 if the subject's social norm matched the chosen principle. Robust standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Social Preference and Norm by chosen Distribution

<i>Social Preference</i>	Chosen Distribution			
	Inequality Aversion	Maximin	Meritocracy	Utilitarianism
Inequality Aversion	11.25%	75.00%	5.00%	8.75%
Maximin	7.14%	73.81%	2.38%	16.67%
Meritocracy	4.17%	45.00%	14.17%	36.67%
Utilitarianism	7.69%	65.38%	7.69%	19.23%
<i>Social Norm</i>				
Inequality Aversion	31.37%	56.86%	1.96%	9.80%
Maximin	3.55%	78.72%	4.96%	12.77%
Meritocracy	0.00%	35.71%	32.14%	32.14%
Utilitarianism	0.00%	39.19%	12.16%	48.65%

norm is somewhat larger than the proportion of respondents in the Average Income Test (see table 6).

C.4.3 Main results

Table 11 reports the results of logistic regressions with individual distribution choices as the outcome variables for respondents in the Motivation Test. This test was conducted with only the Non-Veil of Ignorance treatment. These regression results are also directly comparable to table 4 in the main text. Social norms are again a highly significant predictor of distribution

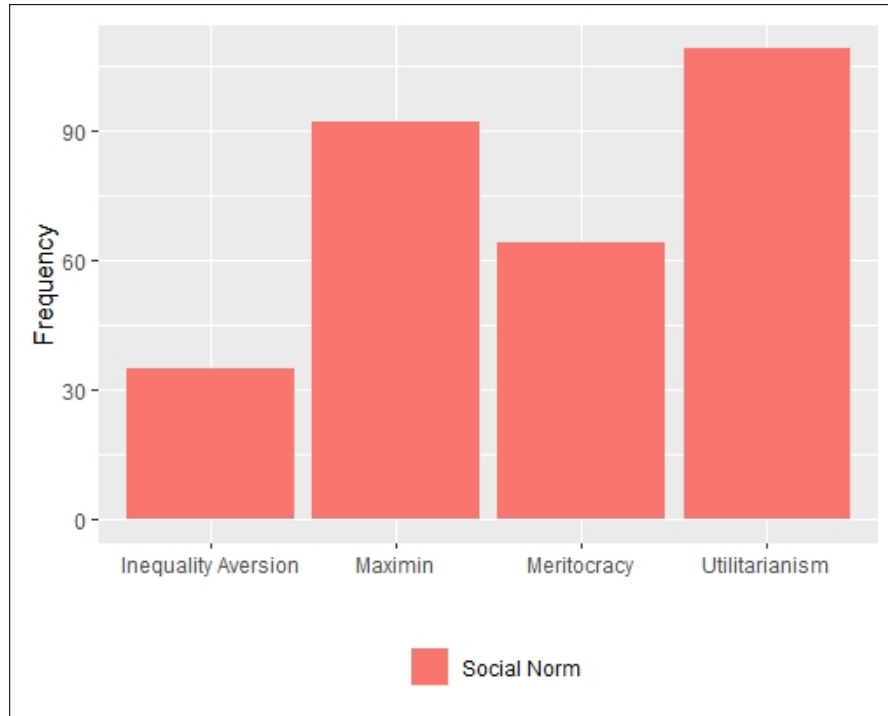
Table 7: Logistic regressions of distributive choices for Average Income Test

	Impartial Spectator Treatment			
	Inequality Aversion	Choice of Distribution		Utilitarianism
		Maximin	Meritocracy	
Social Preference	0.598 (0.529)	0.468 (0.406)	1.372*** (0.516)	-0.296 (0.423)
Social Norm	3.704*** (0.796)	1.770*** (0.302)	2.209*** (0.612)	2.034*** (0.325)
Sample				
<i>United Kingdom</i>	-1.453* (0.769)	0.427 (0.386)	-0.652 (0.611)	0.146 (0.438)
<i>United States</i>	-0.737 (0.798)	0.277 (0.479)	-0.746 (0.681)	0.160 (0.542)
Quiz Performance	-0.066 (0.264)	-0.076 (0.132)	0.020 (0.228)	0.080 (0.146)
Income	-0.346 (0.284)	0.149 (0.119)	0.162 (0.156)	-0.147 (0.146)
Female	-0.540 (0.634)	0.389 (0.304)	0.379 (0.528)	-0.594* (0.346)
Left-Right	-0.058 (0.379)	-0.089 (0.203)	0.767** (0.368)	-0.191 (0.241)
Age	0.338 (0.273)	-0.139 (0.137)	-0.313 (0.250)	0.316* (0.167)
Risk seeking	-0.013 (0.127)	-0.130 (0.080)	0.058 (0.152)	0.140 (0.089)
Student	-0.343 (0.584)	0.157 (0.350)	-0.416 (0.720)	0.104 (0.439)
Economics	1.320* (0.700)	-0.415 (0.338)	0.069 (0.561)	0.190 (0.395)
Constant	-3.335** (1.567)	0.552 (0.991)	-6.208*** (1.748)	-2.559** (1.145)
Observations	271	271	271	271
Pseudo R-squared	0.349	0.145	0.176	0.160

Notes: Estimates come from a logistic regression. Social Preference is a binary variable equal to 1 if the subject's choice of principle matched the distribution. Social Norm is a binary variable equal to 1 if the subject's social norm matched the distribution choice. The reference category for the sample variables is Western Europe. Quiz performance ranges from 0 to 5 depending on how many questions the subject answered correctly. A higher value on the left-right variable indicates a more left-wing orientation on economic policy. Risk preferences are self-reported on a scale from 0 to 10 with 10 being the most risk-seeking option. Student is a binary variable equal to 1 if the subject is currently studying towards a degree and Economics is a binary variable equal to 1 if the subject has ever studied a course on Economics at University. Robust standard errors are presented in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

choices while social preferences only matter for the distribution choices of respondents who chose the Meritocratic or Maximin distribution with much smaller coefficients. The social norm coefficients are similar to those of the main regression results. As this test included only the Non-Veil of Ignorance treatment we could also include a selfishness variable. Contrary to our main results, selfishness is negatively associated with choosing the meritocratic and utilitarian distribution, yet positively associated with choosing the Maximin distribution.

Figure 5: Distribution of perceived Social Norm



C.4.4 Preference-following in perceived Social Norm

Table 12 reports individual characteristics for respondents who chose a perceived social norm which is equivalent to their stated social preference. A stronger social identification is a weakly significant predictor of having a social preference that is equivalent to the perceived social norm.

C.5 Second Principle Test

C.5.1 Second Principle Distribution

Out of the 201 subjects included in the second principle test 113 indicated that they would take a second principle into consideration when deciding on how to distribute income in the group.

Table 13 reports the chosen second principle by first principle. The first thing to note is that Maximin is not the most chosen second choice of any of the first principles. In fact, it is the least chosen second option. We additionally find that subjects are on average significantly ($p=0.002$) more confident in their first choice of principle (average of 7.325 on a 10-point scale) as opposed to their second choice (average of 6.673).

Table 8: Logistic regressions of individual characteristics by subject group

	Non-Veil of Ignorance Treatment			
	Subject Group			
	Social Preference Followers		Norm Followers	
Ambiguity preference	0.035 (0.077)	-0.007 (0.083)	-0.198*** (0.075)	-0.200** (0.081)
Confidence	0.079** (0.038)	0.071* (0.042)	0.000 (0.035)	0.011 (0.039)
Identity	0.025 (0.079)	0.049 (0.088)	-0.103 (0.071)	-0.127 (0.078)
Identity group				
Ethnicity	0.478 (0.387)	0.747* (0.428)	-0.677* (0.347)	-0.619 (0.390)
Nationality	0.441 (0.301)	0.630 (0.339)	-0.487* (0.267)	-0.530* (0.288)
Occupation	0.367 (0.318)	0.487 (0.356)	-0.479* (0.285)	-0.503 (0.308)
Race	0.339 (0.524)	0.796 (0.564)	-1.352*** (0.457)	-1.269** (0.542)
Religion	0.580 (0.505)	0.653 (0.548)	-0.035 (0.475)	0.077 (0.502)
Self-deception 1	0.004 (0.049)	0.005 (0.053)	-0.050 (0.043)	-0.028 (0.048)
Self-deception 2	0.019 (0.043)	-0.020 (0.046)	-0.023 (0.038)	0.013 (0.042)
Constant	-2.344*** (0.619)	-2.474*** (0.892)	2.300*** (0.569)	2.651*** (0.798)
Individual Controls		✓		✓
Observations	971	859	971	859
Pseudo R-squared	0.006	0.020	0.017	0.041

Notes: Estimates come from a logistic regression. The outcome variable 'Social Preference Followers' is equal to 1 if the subject followed their social preference in the distribution choice and 0 otherwise. The outcome variable "Norm Followers" is equal to 1 if the subject followed the perceived social norm in their distribution choice and 0 otherwise. Ambiguity preference ranges from 0 to 7 (with a higher value indicating more ambiguity seeking preferences) and is a standardized scale based on the ambiguity preference survey module developed by [Cavatorta and Schröder \(2019\)](#). Confidence is measured from 1 to 10 and a higher value indicates more confidence in the chosen principle. Identity ranges from 1 to 4 with a higher value indicating a higher level of identity. This variable was measured using the module developed by [Kuo and Margalit \(2012\)](#). 'Other' is the reference group for identity group. Self-deception 1 ranges from 1 to 7 with a lower value indicating more self-deception. Self-deception 2 ranges from 1 to 7 with a higher value indicating more tolerance for deception. *** p<0.01, ** p<0.05, * p<0.1.

C.6 Distribution Test

C.6.1 Assumed distribution by chosen principle

Out of the 200 subjects included in the distribution test, 89 correctly identified the distribution associated with their chosen principle. Table 14 reports the distribution subjects assumed to represent the chosen principle by chosen principle. Subjects who chose Maximin as their principle were by far the best at identifying the distribution corresponding to their principle (77.14% correctly identified the distribution). Out of those who chose Meritocracy as their principle (which is the majority of subjects in our main waves), only 8% confused

Table 9: Linear Regression of Confidence in Principle

	Confidence in Principle	
Ambiguity preference	-0.001 (0.003)	-0.001 (0.003)
Identity	0.175** (0.070)	0.162** (0.074)
Identity group		
Ethnicity	-0.506 (0.318)	-0.114 (0.327)
Nationality	-0.451* (0.236)	-0.256 (0.257)
Occupation	-0.497* (0.260)	-0.254 (0.278)
Race	-0.219 (0.422)	-0.046 (0.471)
Religion	-0.684* (0.395)	-0.178 (0.407)
Self-deception 1	0.039 (0.042)	0.022 (0.043)
Self-deception 2	0.051 (0.035)	0.020 (0.038)
Constant	6.683*** (0.815)	6.683*** (0.951)
Individual Controls		✓
Observations	971	859
Pseudo R-squared	0.018	0.083

Notes: Estimates come from a linear regression. The outcome variable 'Confidence in Principle' is measured from 1 to 10 and a higher value indicates more confidence in the chosen principle. Ambiguity preference ranges from 0 to 7 (with a higher value indicating more ambiguity seeking preferences) and is a standardized scale based on the ambiguity preference survey module developed by [Cavatorta and Schröder \(2019\)](#). Identity ranges from 1 to 4 with a higher value indicating a higher level of identity. This variable was measured using the module developed by [Kuo and Margalit \(2012\)](#). 'Other' is the reference group for identity group. Self-deception 1 ranges from 1 to 7 with a lower value indicating more self-deception. Self-deception 2 ranges from 1 to 7 with a higher value indicating more tolerance for deception. *** p<0.01, ** p<0.05, * p<0.1.

the Maximin distribution with the meritocratic distribution. Most of those subjects thought the utilitarian distribution to be the meritocratic one. This emphasises the robustness of our main result, as meritocrats did not move towards Maximin out of confusion.

Table 10: Social Preference and Norm by chosen Distribution

<i>Social Preference</i>	Chosen Distribution			
	Inequality Aversion	Maximin	Meritocracy	Utilitarianism
Inequality Aversion	18.63%	55.23%	11.11%	15.03%
Maximin	11.94%	63.43%	5.97%	18.66%
Meritocracy	12.61%	30.95%	19.48%	36.96%
Utilitarianism	18.40%	54.25%	11.79%	15.57%
<i>Social Norm</i>				
Inequality Aversion	58.06%	24.19%	5.65%	12.10%
Maximin	10.61%	73.86%	6.44%	9.09%
Meritocracy	7.27%	20.00%	44.55%	28.18%
Utilitarianism	8.37%	14.64%	18.83%	58.16%

C.6.2 Main analysis excluding subjects with inequality aversion as a first principle

Out of those subjects who chose inequality aversion as their principle, 52.54% confused the Maximin distribution with the inequality averse distribution. Given that this probably explains some of the movement towards Maximin in the distribution choice, we repeated our main analysis excluding those who chose inequality aversion as their principle in table 15. It is evident from the results reported in the table that excluding those with inequality aversion as their principle does not affect our main result - social norms are still significantly better predictors of distribution choices than social preferences. This result holds even when we only look at subjects in the non-veil of ignorance treatment and control for selfishness.

Table 11: Logistic regressions of distributive choices for Motivation Test

	Non-Veil of Ignorance Treatment			
	Inequality Aversion	Choice of Distribution		
		Maximin	Meritocracy	Utilitarianism
Social Preference	-0.002 (0.219)	0.530** (0.251)	0.380* (0.227)	-0.235 (0.246)
Social Norm	2.388*** (0.227)	2.613*** (0.196)	1.574*** (0.275)	2.059*** (0.207)
Selfishness	0.326 (0.206)	2.247*** (0.206)	-3.694*** (0.716)	-3.027*** (0.391)
Sample				
<i>United Kingdom</i>	0.033 (0.225)	0.083 (0.211)	-0.311 (0.262)	0.231 (0.230)
<i>United States</i>	0.279 (0.505)	-0.268 (0.652)	-0.121 (0.510)	0.176 (0.595)
Quiz Performance	-0.062 (0.084)	0.070 (0.083)	-0.121 (0.112)	0.014 (0.090)
Income	-0.000 (0.107)	-0.141 (0.118)	0.065 (0.111)	0.074 (0.104)
Female	0.014 (0.211)	-0.081 (0.207)	0.143 (0.258)	0.017 (0.222)
Left-Right	0.041 (0.140)	-0.082 (0.119)	-0.074 (0.147)	0.065 (0.145)
Age	0.104 (0.107)	-0.038 (0.100)	0.052 (0.124)	-0.116 (0.110)
Risk seeking	-0.025 (0.051)	-0.091* (0.048)	0.162*** (0.058)	0.032 (0.051)
Student	0.167 (0.262)	-0.144 (0.238)	0.467* (0.283)	-0.330 (0.241)
Economics	0.041 (0.264)	0.118 (0.248)	-0.134 (0.296)	0.014 (0.264)
Constant	-2.505*** (0.849)	-1.561** (0.703)	-2.398*** (0.900)	-1.604* (0.832)
Observations	886	886	886	886
Pseudo R-squared	0.157	0.394	0.254	0.321

Notes: Estimates come from a logistic regression. Social Preference is a binary variable equal to 1 if the subject's choice of principle matched the distribution. Social Norm is a binary variable equal to 1 if the subject's social norm matched the distribution choice. Selfishness is a binary variable equal to 1 if the subject chose the distribution that maximises the payoff of the quintile they were placed in based on their example quiz answers. The reference category for the sample variables is Western Europe. Quiz performance ranges from 0 to 5 depending on how many questions the subject answered correctly. A higher value on the left-right variable indicates a more left-wing orientation on economic policy. Risk preferences are self-reported on a scale from 0 to 10 with 10 being the most risk-seeking option. Student is a binary variable equal to 1 if the subject is currently studying towards a degree and Economics is a binary variable equal to 1 if the subject has ever studied a course on Economics at University. Robust standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Logistic Regression of Preference-following in perceived Social Norm

	Preference-followers in perceived Norm	
Ambiguity preference	0.108 (0.084)	0.088 (0.091)
Identity	0.156* (0.084)	0.165* (0.095)
Identity group		
Ethnicity	0.402 (0.365)	0.697* (0.406)
Nationality	-0.056 (0.282)	0.121 (0.318)
Occupation	-0.056 (0.305)	0.072 (0.342)
Race	0.588 (0.479)	0.762 (0.541)
Religion	0.457 (0.480)	0.403 (0.529)
Self-deception 1	-0.029 (0.050)	-0.036 (0.055)
Self-deception 2	0.021 (0.045)	0.003 (0.049)
Constant	-2.558*** (0.619)	-2.257** (0.881)
Individual Controls		✓
Observations	971	859
Pseudo R-squared	0.017	0.027

Notes: Estimates come from a linear regression. The outcome variable 'Preference-followers in perceive Norm' is equal to 1 if the subject's perceived social norm is equivalent to their stated social preference and 0 otherwise. Ambiguity preference ranges from 0 to 7 (with a higher value indicating more ambiguity seeking preferences) and is a standardized scale based on the ambiguity preference survey module developed by [Cavatorta and Schröder \(2019\)](#). Confidence is measured from 1 to 10 and a higher value indicates more confidence in the chosen principle. Identity ranges from 1 to 4 with a higher value indicating a higher level of identity. This variable was measured using the module developed by [Kuo and Margalit \(2012\)](#). 'Other' is the reference group for identity group. Self-deception 1 ranges from 1 to 7 with a lower value indicating more self-deception. Self-deception 2 ranges from 1 to 7 with a higher value indicating more tolerance for deception. *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Social Preference and Norm by chosen Distribution

First Principle				
<i>Second Principle</i>	Inequality Aversion	Maximin	Meritocracy	Utilitarianism
Inequality Aversion	0.00%	58.62%	51.35%	70.59%
Maximin	26.67%	0.00%	21.62%	5.88%
Meritocracy	33.33%	27.59%	0.00%	23.53%
Utilitarianism	40.00%	13.79%	27.03%	0.00%
Total	100%	100%	100%	100%

Table 14: Assumed distribution by chosen principle

Chosen Principle				
<i>Distribution</i>	Inequality Aversion	Maximin	Meritocracy	Utilitarianism
Inequality Aversion	37.29%	5.71%	4.00%	22.58%
Maximin	52.54%	77.14%	8.00%	25.81%
Meritocracy	3.39%	5.71%	33.33%	3.23%
Utilitarianism	6.78%	11.43%	54.67%	48.39%
Total	100%	100%	100%	100%

Table 15: Logistic regressions of distributive choices (excluding inequality aversion preference-holders)

	All Treatments				Non-Veil of Ignorance Treatment			
	Inequality Aversion	Choice of Distribution			Inequality Aversion	Choice of Distribution		
		Maximin	Meritocracy	Utilitarianism		Maximin	Meritocracy	Utilitarianism
Social Preference		1.075*** (0.153)	0.937*** (0.181)	-0.956*** (0.154)		0.792*** (0.268)	0.534* (0.322)	-0.401 (0.245)
Social Norm	2.601*** (0.176)	2.140*** (0.120)	1.874*** (0.182)	2.013*** (0.127)	2.647*** (0.306)	2.297*** (0.221)	1.644*** (0.320)	1.924*** (0.221)
Selfishness					0.454 (0.284)	-0.631*** (0.215)	0.501* (0.269)	0.024 (0.206)
Treatments								
<i>Veil of Ignorance</i>	1.034*** (0.222)	-0.046 (0.140)	-0.080 (0.195)	-0.255* (0.151)				
<i>Non-Veil of Ignorance</i>	0.866*** (0.230)	-0.168 (0.144)	-0.181 (0.201)	-0.005 (0.155)				
Sample								
<i>United Kingdom</i>	-0.068 (0.230)	0.166 (0.159)	0.283 (0.241)	-0.210 (0.161)	-0.159 (0.509)	0.614* (0.305)	-0.026 (0.424)	-0.160 (0.267)
<i>United States</i>	0.008 (0.302)	-0.157 (0.193)	0.471* (0.271)	-0.100 (0.194)	-0.159 (0.509)	-0.235 (0.360)	0.850* (0.465)	-0.152 (0.326)
Constant	-4.094*** (0.670)	-2.360*** (0.455)	-2.974*** (0.660)	-0.034 (0.472)	-3.073*** (1.148)	-1.496* (0.773)	-5.024*** (1.193)	-0.195 (0.862)
Session Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1,645	1,645	1,645	1,645	530	530	530	530
Pseudo R-squared	0.224	0.214	0.134	0.192	0.220	0.230	0.138	0.182

Notes: Estimates come from a logistic regression. Social Preference is a binary variable equal to 1 if the subject's choice of principle matched the distribution. Social Norm is a binary variable equal to 1 if the subject's social norm matched the distribution choice. Selfishness is a binary variable equal to 1 if the subject chose the distribution that maximises the payoff of the quintile they were placed in based on their example quiz answers. The reference category for the treatment variables is the Impartial Spectator treatment. The reference category for the sample variables is Western Europe. Robust standard errors are presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

C.7 Wording Test

C.7.1 Assumed distribution by chosen principle

Given the alternative wording of the maximin and inequality aversion statements used in our wording test, we first check whether the proportion of subjects correctly identifying the corresponding distribution has changed. Out of the 222 subjects included in the wording test, 88 correctly identified the distribution associated with their chosen principle. This is a significantly smaller proportion than subjects who correctly identified the distribution associated with their chosen principle when we used the original wording (39.64% compared to 44.50%). This finding therefore supports the use of our original statements in our main analysis. Table 16 reports the distribution subjects assumed to represent the chosen principle by chosen principle. The percentages are strikingly similar to those reported in table 14 of this appendix. Importantly, however, the proportion of respondents who correctly identified maximin and inequality aversion, the two principles for which the wording changed, decreased. In fact, the percentage of subjects correctly identifying inequality aversion as the distribution corresponding to their chosen principle decreased from just over 37% to about 29%.

Table 16: Assumed distribution by chosen principle - alternative wording

<i>Distribution</i>	Chosen Principle			
	Inequality Aversion	Maximin	Meritocracy	Utilitarianism
Inequality Aversion	28.99%	3.57%	9.41%	19.51%
Maximin	53.62%	75.00%	11.76%	26.83%
Meritocracy	10.14%	7.14%	31.76%	2.44%
Utilitarianism	7.25%	14.29%	47.06%	51.22%
Total	100%	100%	100%	100%

C.7.2 Main results

Table 17 reports the results of logistic regressions with individual distribution choices as the outcome variables for respondents in the Wording Test. This test was conducted with only the Impartial Spectator treatment. These regression results are directly comparable to table 4 in the main text. Despite the small sample size of the robustness check and the lower proportion of subjects who correctly identified the distribution corresponding to their principle, the main results are strikingly robust. Social norms are a consistent and

highly significant predictor of distribution choices while social preferences are mostly not. Only those choosing the meritocratic distribution are significantly affected by their social preference. The social norm coefficients are again similar to those of the main regression results.

Table 17: Logistic regressions of distributive choices for Wording Test

	Impartial Spectator Treatment			
	Inequality Aversion	Choice of Distribution		Utilitarianism
		Maximin	Meritocracy	
Social Preference	-0.223 (0.658)	0.501 (0.446)	1.412*** (0.509)	-0.608 (0.635)
Social Norm	3.250*** (0.679)	1.661*** (0.343)	2.727*** (0.611)	2.000*** (0.398)
Sample				
<i>United Kingdom</i>	1.241 (1.059)	-0.743 (0.451)	0.608 (0.635)	0.038 (0.553)
<i>United States</i>	1.306 (0.998)	-0.615 (0.506)	-0.122 (0.575)	0.140 (0.564)
Income	-0.191 (0.315)	-0.164 (0.159)	0.168 (0.256)	0.150 (0.161)
Female	0.277 (0.620)	0.476 (0.353)	-0.191 (0.538)	-0.572 (0.442)
Left-Right	0.119 (0.518)	0.076 (0.221)	0.283 (0.320)	-0.234 (0.268)
Age	-0.176 (0.299)	-0.199 (0.199)	0.004 (0.317)	0.316 (0.248)
Risk seeking	-0.034 (0.111)	0.018 (0.086)	-0.204* (0.122)	0.107 (0.102)
Student	0.063 (0.854)	-0.753* (0.410)	0.885 (0.641)	0.288 (0.534)
Economics	-0.857 (0.822)	-0.159 (0.392)	1.376** (0.532)	-0.237 (0.470)
Constant	-3.276 (2.595)	0.169 (1.090)	-4.415*** (1.614)	-2.491* (1.281)
Observations	187	187	187	187
Pseudo R-squared	0.271	0.141	0.308	0.167

Notes: Estimates come from a logistic regression. Social Preference is a binary variable equal to 1 if the subject's choice of principle matched the distribution. Social Norm is a binary variable equal to 1 if the subject's social norm matched the distribution choice. The reference category for the sample variables is Western Europe. A higher value on the left-right variable indicates a more left-wing orientation on economic policy. Risk preferences are self-reported on a scale from 0 to 10 with 10 being the most risk-seeking option. Student is a binary variable equal to 1 if the subject is currently studying towards a degree and Economics is a binary variable equal to 1 if the subject has ever studied a course on Economics at University. Robust standard errors are presented in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C.8 Order Test

C.8.1 Preference- and Norm-following by chosen Distribution

Table 18 reports the chosen distribution by stated social preference and social norm for respondents in the Order Test. The pattern visible in table 18 is similar to the results of the main experiment: Social Norms are more closely related to distribution choices than

social preferences. The percentage of norm followers is especially high for Maximin with over 72% of respondents who chose the Maximin distribution following their perceived social norm. Interestingly, given this reversed order of decisions, the percentage of those who chose the inequality averse and utilitarian distributions and follow their perceived social norm decreased while the opposite is the case for those who chose the meritocratic distribution, compared to the results of our main waves.

Table 18: Social Preference and Norm by chosen Distribution

<i>Social Preference</i>	Chosen Distribution			
	Inequality Aversion	Maximin	Meritocracy	Utilitarianism
Inequality Aversion	16.13%	62.90%	11.29%	9.68%
Maximin	3.33%	53.33%	13.33%	30.00%
Meritocracy	5.88%	16.47%	42.35%	35.29%
Utilitarianism	12.20%	48.78%	9.76%	29.27%
<i>Social Norm</i>				
Inequality Aversion	40.63%	28.13%	12.50%	18.75%
Maximin	5.56%	72.22%	13.89%	8.33%
Meritocracy	8.57%	17.14%	51.43%	22.86%
Utilitarianism	1.27%	27.85%	24.05%	46.84%

C.8.2 Main results

Table 19 reports the results of logistic regressions with individual distribution choices as the outcome variables for respondents in the order test. This test was conducted with only the Impartial Spectator treatment. These regression results are directly comparable to table 4 in the main text. Similar to all previous robustness checks, the main results hold again. Despite the small sample size of this robustness check, social norms are a highly significant predictor of distribution choices for all principles. The social norm coefficients are again similar to those of the main regression results, although, given the smaller sample size, there is some more variation. For example, the coefficient for inequality aversion is over twice as large as the one reported in table 4 of the main text. Social preferences are also significant for all distribution choices except utilitarianism, although except for meritocracy, the coefficients are smaller than those for social norms.

Table 19: Logistic regressions of distributive choices for Order Test

	Impartial Spectator Treatment			
	Inequality Aversion	Choice of Distribution		
		Maximin	Meritocracy	Utilitarianism
Social Preference	2.607*** (0.904)	1.178** (0.528)	2.200*** (0.477)	0.634 (0.433)
Social Norm	5.544*** (0.951)	2.365*** (0.366)	1.961*** (0.509)	1.977*** (0.408)
Sample				
<i>United Kingdom</i>	-2.114* (1.125)	-0.201 (0.452)	0.322 (0.514)	0.360 (0.503)
<i>United States</i>	-0.981 (0.843)	-0.586 (0.532)	0.515 (0.554)	0.312 (0.496)
Income	-0.323 (0.427)	-0.103 (0.155)	-0.041 (0.178)	0.178 (0.155)
Female	2.582** (1.180)	-0.298 (0.323)	0.247 (0.439)	-0.276 (0.406)
Left-Right	0.243 (0.562)	0.176 (0.202)	-0.007 (0.229)	-0.160 (0.218)
Age	0.006 (0.566)	-0.074 (0.182)	0.469** (0.216)	-0.196 (0.218)
Risk seeking	0.295 (0.333)	0.073 (0.092)	-0.105 (0.105)	-0.113 (0.087)
Student	-2.668** (1.328)	-0.813 (0.497)	1.218** (0.501)	0.314 (0.477)
Economics	1.225 (0.858)	-0.120 (0.419)	0.094 (0.494)	-0.067 (0.424)
Constant	-7.531 (5.519)	-1.351 (1.151)	-4.068*** (1.436)	-0.799 (1.132)
Observations	196	196	196	196
Pseudo R-squared	0.464	0.217	0.242	0.155

Notes: Estimates come from a logistic regression. Social Preference is a binary variable equal to 1 if the subject's choice of principle matched the distribution. Social Norm is a binary variable equal to 1 if the subject's social norm matched the distribution choice. The reference category for the sample variables is Western Europe. A higher value on the left-right variable indicates a more left-wing orientation on economic policy. Risk preferences are self-reported on a scale from 0 to 10 with 10 being the most risk-seeking option. Student is a binary variable equal to 1 if the subject is currently studying towards a degree and Economics is a binary variable equal to 1 if the subject has ever studied a course on Economics at University. Robust standard errors are presented in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

D DESCRIPTION OF VARIABLES

Principle. Categorical variable capturing the principle selected by subject i .

- 1: Utilitarianism
- 2: Meritocracy
- 3: Maximin
- 4: Inequality Aversion.

Distribution. Categorical variable capturing the distribution selected by subject i .

- 1: Utilitarianism
- 2: Meritocracy

- 3: Maximin
- 4: Inequality Aversion.

Principle Norm. Categorical variable capturing the perceived social norm for the principle choice selected by subject i.

- 1: Utilitarianism
- 2: Meritocracy
- 3: Maximin
- 4: Inequality Aversion.

Distribution Norm. Categorical variable capturing the perceived social norm for the distribution choice selected by subject i.

- 1: Utilitarianism
- 2: Meritocracy
- 3: Maximin
- 4: Inequality Aversion.

Treatment. Categorical variable capturing the treatment subject i is assigned to.

- 1: Impartial Spectator
- 2: Non-Veil of Ignorance
- 3: Veil of Ignorance

Gender. Binary variable coded as 1 if subject i indicated to be female, 0 if subject i indicated to be male. Subjects who indicated "other" or "prefer not to say" were coded as missing values (n=22).

Age. Categorical variable capturing the age bracket of subject i.

- 1: 18-20 years old
- 2: 21-29 years old
- 3: 30-39 years old
- 4: 40-49 years old
- 5: 50-59 years old
- 6: 60 years or older

Student. Binary variable coded as 1 if subject i is currently in full-time education, 0 otherwise.

Economics. Binary variable coded as 1 if subject i indicated that they have taken a module in economics or a related subject at University. A value of 0 indicates that subject i either has not taken a module in economics or has never attended higher education.

Left-Right. Categorical variable capturing how much subject i agrees with the statement: "On economic policy matters, there is a role for the government".

- 1: Strongly Disagree
- 2: Disagree
- 3: Neither Agree nor Disagree
- 4: Agree
- 5: Strongly agree

Risk preferences. Variable capturing subject i 's willingness to take risks on a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means "very willing to take risks".

Income. Categorical variable capturing the income bracket of subject i . Values are stated in Pound Sterling (£) for subjects from the UK, in US Dollars (\$) for subjects from the US and in Euros (€) for subjects from Western Europe.

- 1: Less than 20,000
- 2: 20,000 to 34,999
- 3: 35,000 to 49,999
- 4: 50,000 to 74,999
- 5: 75,000 to 99,999
- 6: Over 100,000

Sample. Categorical variable indicating whether subject i is a resident in the US, UK or Western Europe.

- 1: Europe
- 2: United Kingdom
- 3: United States

Quiz Performance. Variable ranging from 0 to 5, capturing the number of questions subject i correctly answered in the main Quiz.

Example Quiz Performance. Variable ranging from 0 to 2, capturing the number of questions subject i correctly answered in the example quiz of the Non-Veil of Ignorance treatment.

Study. Variable indicating whether subject i was part of the first wave of the main study in November 2019 or the second wave in December 2019.

Principle Following. Binary variable coded as 1 if subject i 's chosen distribution is equal to their chosen principle.

Norm Following. Binary variable coded as 1 if subject i 's chosen distribution is equal to their perceived social norm in the distribution choice.

Principle Following in perceived Social Norm. Binary variable coded as 1 if subject i 's perceived social norm in the distribution choice is equal to their chosen principle.

Norm Following in Principle. Binary variable coded as 1 if subject i 's chosen principle is equal to their perceived social norm in the principle choice.

Selfish. Binary variable coded as 1 if subject i 's chosen distribution is the distribution which maximises the income of their predicted quintile position from the example quiz in the Non-Veil of Ignorance treatment.

Decision Group. Categorical variable indicating whether subject i is a norm-follower, principle-follower or selfish in the Non-Veil of Ignorance treatment. Subjects that are both, norm- and principle-followers, are coded as principle-followers. Subjects that are both, norm-followers and selfish, are coded as selfish. Subjects that are both, principle-followers and selfish, are coded as selfish. This coding is used to ensure the most robust test of our hypotheses.

- 1: Norm-Following
- 2: Principle-Following
- 3: Selfish

Confidence in Principle. Variable capturing subject i 's confidence in their chosen principle on a scale from 0 to 10, where 0 means "not confident at all" and a 10 means "Very confident".

Identity. Variable capturing subject i 's social identification with a self-defined reference group, ranging from 1 to 4 with 1 indicating "Not strong at all" and 4 indicating "Very strong" social identity.

Identity Group. Categorical variable capturing the group subject i most identifies with. This variable is also used as the reference group for the Identity variable.

- 1: Your ethnicity
- 2: Your nationality
- 3: Your occupation
- 4: Your race
- 5: Your religion
- 6: Other

Self-Deception 1. Variable capturing subject i 's self-deception measured as the level of agreement with the statement "It is okay to lie sometimes", ranging from 1 to 7 whereby 1 means "Strongly agree" and 7 means "Strongly disagree".

Self-Deception 2. Variable capturing subject i 's self-deception measured as their response to the statement "There is a big debate in psychology over whether deception in experiments should be permitted. What do you think?", ranging from 1 to 7 whereby 1 means "Never" and 7 means "Whenever it helps science".

Ambiguity preference. Variable capturing subject i 's preference for ambiguity ranging from 0 to 7 with 0 indicating ambiguity aversion and 7 ambiguity seeking preferences.

E SURVEY INSTRUMENT

All values below are given in Pound Sterling (£). This was changed to US Dollars (\$) and Euros (€) depending on respondents' country of residence. All options in decisions 1-4 were presented in randomized order during the survey experiment. Distribution options in decisions 3 and 4 were presented as separate tables.

E.1 Impartial Spectator Treatment

Background

A group of people are asked to do a quiz and their answers generate income. We rank their performance from the bottom 20% of performers to the top 20% in the table below and give the average income generated for a person in each 20% performance band. For example, this shows someone who performs in the middle band (the 3rd 20%) generates an income of £40 on average. Please click on the arrow below to proceed.

<i>Performance Level</i>	<i>Average Income</i>
Bottom 20% of performers	£20
2nd 20%	£30
3rd 20%	£40
4th 20%	£70
5th 20%	£110

Decision 1

Which of the following statements best describes how you think income should be distributed in this group? Please note, you are not part of this group.

- Inequalities are only justifiable if they improve the position of the least well-off group in society.
- Inequalities should be minimized.
- Individual income should be based exclusively on his/her ability and talents.
- Income should be distributed to maximize the average income in society.

Decision 2

All the participants of the study are now asked to select a statement. Each of you will be rewarded with a bonus payment of 50p if you select the statement chosen by most of the participants.

- Inequalities are only justifiable if they improve the position of the least well-off group in society.
- Inequalities should be minimized.
- Individual income should be based exclusively on his/her ability and talents.

- Income should be distributed to maximize the average income in society.

Decision 3

Below you can see four options for distributing the income generated by the quiz. It shows for each option how much a performer in each 20% band will receive. For example, a performer in the bottom 20% can either receive £20, £30 or £40 depending on the distribution. As mentioned earlier, performance on the quiz generates income for this group on average as in the table below:

<i>Performance Level</i>	<i>Average Income</i>
Bottom 20% of performers	£20
2nd 20%	£30
3rd 20%	£40
4th 20%	£70
5th 20%	£110

Which distribution option would you choose for this group? Please note, you are not part of this group.

Average Income				
<i>Performance Level</i>	<i>Inequality Aversion</i>	<i>Maximin</i>	<i>Meritocracy</i>	<i>Utilitarianism</i>
Bottom 20%	£30	£40	£20	£20
2nd 20%	£60	£40	£30	£30
3rd 20%	£60	£50	£40	£50
4th 20%	£60	£60	£70	£70
5th 20%	£60	£80	£110	£110
Total	£270	£270	£270	£280

Decision 4

All the participants of the study are now asked to select a distribution. Each of you will be rewarded with a bonus payment of 50p if you select the distribution chosen by most of the participants.

Average Income				
<i>Performance Level</i>	<i>Inequality Aversion</i>	<i>Maximin</i>	<i>Meritocracy</i>	<i>Utilitarianism</i>
Bottom 20%	£30	£40	£20	£20
2nd 20%	£60	£40	£30	£30
3rd 20%	£60	£50	£40	£50
4th 20%	£60	£60	£70	£70
5th 20%	£60	£80	£110	£110
Total	£270	£270	£270	£280

Quiz Introduction

You will now take part in the previously mentioned quiz which is the final part of this study. You will have 30 seconds to answer as many questions as possible. For your participation in the quiz you will receive an additional bonus payment of 50ct after completing the study. However, how well you perform on the quiz does not influence the size of this bonus payment.

E.2 Veil of Ignorance Treatment

Background

People in a group that you belong to are asked to do a quiz and their answers generate income. We rank performance from the bottom 20% of performers to the top 20% in the table below and give the average income generated for a person in each 20% performance band. For example, the table below shows someone who performs in the middle band (the 3rd 20%) generates an income of £40 on average. In the following, you will participate in the above mentioned quiz and your performance will affect the bonus payment you will receive after completing the study. Please click on the arrow below to continue.

<i>Performance Level</i>	<i>Average Income</i>
Bottom 20% of performers	£20
2nd 20%	£30
3rd 20%	£40
4th 20%	£70
5th 20%	£110

Decision 1

Which of the following statements best describes how you think income should be distributed in your group?

- Inequalities are only justifiable if they improve the position of the least well-off group in society.
- Inequalities should be minimized.
- Individual income should be based exclusively on his/her ability and talents.
- Income should be distributed to maximize the average income in society.

Decision 2

All the participants in your group are now asked to select a statement. Each of you will be rewarded with a bonus payment of 50p if you select the statement chosen by most of the members of your group.

- Inequalities are only justifiable if they improve the position of the least well-off group in society.
- Inequalities should be minimized.

- Individual income should be based exclusively on his/her ability and talents.
- Income should be distributed to maximize the average income in society.

Decision 3

Below you can see four options for distributing the income generated by the quiz. It shows for each option how much a performer in each 20% band will receive. For example, a performer in the bottom 20% can either receive £20, £30 or £40 depending on the distribution. As mentioned earlier, performance on the quiz generates income for your group on average as in the table below:

<i>Performance Level</i>	<i>Average Income</i>
Bottom 20% of performers	£20
2nd 20%	£30
3rd 20%	£40
4th 20%	£70
5th 20%	£110

Which distribution option would you like to choose for your group? The distribution you choose will be implemented and affect the bonus payment you can earn through your performance on the quiz. The conversion rate for the bonus payment is £1=1p so if your performance puts you into the top 20% you can receive a bonus payment of 60p-110p depending on the distribution you have chosen.

<i>Average Income</i>				
<i>Performance Level</i>	<i>Inequality Aversion</i>	<i>Maximin</i>	<i>Meritocracy</i>	<i>Utilitarianism</i>
Bottom 20%	£30	£40	£20	£20
2nd 20%	£60	£40	£30	£30
3rd 20%	£60	£50	£40	£50
4th 20%	£60	£60	£70	£70
5th 20%	£60	£80	£110	£110
Total	£270	£270	£270	£280

Decision 4

All the participants in your group are now asked to select a distribution. Each of you will

be rewarded with a bonus payment of 50p if you select the distribution chosen by most of the members of your group.

Average Income				
<i>Performance Level</i>	<i>Inequality Aversion</i>	<i>Maximin</i>	<i>Meritocracy</i>	<i>Utilitarianism</i>
Bottom 20%	£30	£40	£20	£20
2nd 20%	£60	£40	£30	£30
3rd 20%	£60	£50	£40	£50
4th 20%	£60	£60	£70	£70
5th 20%	£60	£80	£110	£110
Total	£270	£270	£270	£280

Quiz Introduction

You will now take part in the previously mentioned quiz which is the final part of this study. You will have 30 seconds to answer as many questions as possible. How well you perform on this quiz compared to the other participants determines in which of the five performance quintiles you will be placed. Your previously chosen distribution and your performance on this quiz therefore influence the bonus payment you will receive after completing the study.

E.3 Non-veil of Ignorance Treatment

Background

People in a group that you belong to are asked to do a quiz and their answers generate income. We rank performance from the bottom 20% of performers to the top 20% in the table below and give the average income generated for a person in each 20% performance band. For example, the table below shows someone who performs in the middle band (the 3rd 20%) generates an income of £40 on average. In the following, you will participate in the above mentioned quiz and your performance will affect the bonus payment you will receive after completing the study. Please click on the arrow below to continue.

<i>Performance Level</i>	<i>Average Income</i>
Bottom 20% of performers	£20
2nd 20%	£30
3rd 20%	£40
4th 20%	£70
5th 20%	£110

Example Quiz

Please answer the following two questions. Based on your answers to these two questions we will predict how well you will perform on the quiz. You have 15 seconds to answer the questions.

- $9 \times 13 =$
- $80/2.5 =$

On the basis of your answer to these questions we predict that you would belong to the top/middle/bottom 20% of performers in the full quiz.

Decision 1

Which of the following statements best describes how you think income should be distributed in your group?

- Inequalities are only justifiable if they improve the position of the least well-off group in society.
- Inequalities should be minimized.

- Individual income should be based exclusively on his/her ability and talents.
- Income should be distributed to maximize the average income in society.

Decision 2

All the participants in your group are now asked to select a statement. Each of you will be rewarded with a bonus payment of 50p if you select the statement chosen by most of the members of your group.

- Inequalities are only justifiable if they improve the position of the least well-off group in society.
- Inequalities should be minimized.
- Individual income should be based exclusively on his/her ability and talents.
- Income should be distributed to maximize the average income in society.

Decision 3

Below you can see four options for distributing the income generated by the quiz. It shows for each option how much a performer in each 20% band will receive. For example, a performer in the bottom 20% can either receive £20, £30 or £40 depending on the distribution. As mentioned earlier, performance on the quiz generates income for your group on average as in the table below:

<i>Performance Level</i>	<i>Average Income</i>
Bottom 20% of performers	£20
2nd 20%	£30
3rd 20%	£40
4th 20%	£70
5th 20%	£110

Which distribution option would you like to choose for your group? The distribution you choose will be implemented and affect the bonus payment you can earn through your performance on the quiz. The conversion rate for the bonus payment is £1=1p so if your performance puts you into the top 20% you can receive a bonus payment of 60p-110p depending on the distribution you have chosen.

Average Income				
<i>Performance Level</i>	<i>Inequality Aversion</i>	<i>Maximin</i>	<i>Meritocracy</i>	<i>Utilitarianism</i>
Bottom 20%	£30	£40	£20	£20
2nd 20%	£60	£40	£30	£30
3rd 20%	£60	£50	£40	£50
4th 20%	£60	£60	£70	£70
5th 20%	£60	£80	£110	£110
Total	£270	£270	£270	£280

Decision 4

All the participants in your group are now asked to select a distribution. Each of you will be rewarded with a bonus payment of 50p if you select the distribution chosen by most of the members of your group.

Average Income				
<i>Performance Level</i>	<i>Inequality Aversion</i>	<i>Maximin</i>	<i>Meritocracy</i>	<i>Utilitarianism</i>
Bottom 20%	£30	£40	£20	£20
2nd 20%	£60	£40	£30	£30
3rd 20%	£60	£50	£40	£50
4th 20%	£60	£60	£70	£70
5th 20%	£60	£80	£110	£110
Total	£270	£270	£270	£280

Quiz Introduction

You will now take part in the previously mentioned quiz which is the final part of this study. You will have 30 seconds to answer as many questions as possible. How well you perform on this quiz compared to the other participants determines in which of the five performance quintiles you will be placed. Your previously chosen distribution and your performance on this quiz therefore influence the bonus payment you will receive after completing the study.

E.4 Quiz

Please answer as many of the below questions as possible.

- $3 + 5 =$
- $8 \times 16 =$
- $(5 \times 8) - 12.2 =$
- $100 \times 10/5 =$
- $40/2.5 =$

E.5 Demographics

Nationality. What is your country of birth?

Gender. What is your gender?

- Female
- Male
- Other
- Prefer not to say

Age. How old are you?

- 18-20
- 21-29
- 30-39
- 40-49
- 50-59
- 60 or older
- Prefer not to say

Student. Are you currently studying towards a degree at University?

- Yes
- No

Economics. Have you ever taken a module on economics or a related subject area at University?

- Yes
- No
- I have never attended higher education

Income. What is your total personal income per year?

- Less than £20,000
- £20,000 to £34,999
- £35,000 to £49,999
- £50,000 to £74,999
- £75,000 to £99,999
- Over £100,000
- Prefer not to say

Risk preferences. Please tell us, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale.

Left-Right. How much do you agree or disagree with the following statement: "On economic policy matters, there is a role for the government"?

- Strongly agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

Rational. Were there any particular reasons for the principles and distributions you chose? Please use the field below to explain your choices.

Feedback. Please let us know in the field below whether you have any feedback regarding the study. Were any of the questions or tasks unclear?