DISCUSSION PAPER SERIES

DP20624

THE ETHICAL MIRROR

Maya Eden and Paolo G. Piacquadio

MACROECONOMICS AND GROWTH AND PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

THE ETHICAL MIRROR

Maya Eden and Paolo G. Piacquadio

Discussion Paper DP20624
Published 08 September 2025
Submitted 29 August 2025

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

THE ETHICAL MIRROR

Abstract

Are people's ethical views informative about what is actually ethical? We show that, when the social preference relation is Paretian, it must closely mirror people's ethical views. In particular, people's aversion to income inequality among strangers places tight bounds on the admissible inequality attitudes of the social welfare function. Our results also suggest a new rationale for paternalism: if people are paternalistic about the choices of others, then the social welfare function must be paternalistic as well.

JEL Classification: D30, D60, D90

Keywords: Paternalism, Pareto, Inequality aversion

Maya Eden - meden@brandeis.edu
University of Zurich and Brandeis University and CEPR

Paolo G. Piacquadio - paolo.piacquadio@unisg.ch
University of St. Gallen, University of Oslo and CESifo Research Network

Acknowledgements
The authors are grateful to Sandro Ambuehl, Benoit Decerf, Piotr Dworczak, Ernst Fehr, Brian Jabarian, Matti Liski, Francois Maniquet, Itai Sher, and Kristen Vamsaeter for comments, in addition to audiences at the Global Priorities Institute in Oxford, the PSE Frontiers of Economics and Philosophy Conference, GRAPE, Namur, and Navarra. This project has received funding from the Economic Policy Research Network and the European Research Council under the European Union's Horizon 2020 research and innovation program ERC Starting Grant VALURED (Grant agreement No.~804104) and from the Swiss National Science Foundation project EVAS (Grant agreement No.~10001768), as well as support by the UZH Research Priority Program ``Equality of Opportunity.''

The Ethical Mirror

Maya Eden[1] *and Paolo G. Piacquadio*[2]

August 29, 2025

Abstract

Are people's ethical views informative about what is actually ethical? We show that, when the social preference relation is Paretian, it must closely mirror people's ethical views. In particular, people's aversion to income inequality among strangers places tight bounds on the admissible inequality attitudes of the social welfare function. Our results also suggest a new rationale for paternalism: if people are paternalistic about the choices of others, then the social welfare function must be paternalistic as well.

Keywords: Paternalism, Pareto principle, inequality aversion.
JEL classification: D30, D60, D90.

---

[1]Department of Economics, University of Zurich, University of Brandeis, and CEPR (maya.eden@econ.uzh.ch).

[2]Department of Economics, University of St. Gallen, University of Oslo, and CESifo Research Network (paolo.piacquadio@unisg.ch).

The authors are grateful to Sandro Ambühl, Benoît Decerf, Piotr Dworczak, Ernst Fehr, Brian Jabarian, Matti Liski, Francois Maniquet, Itai Sher, and Kristen Vamsaeter for comments, in addition to audiences at the Global Priorities Institute in Oxford, the PSE Frontiers of Economics and Philosophy Conference, GRAPE, Namur, and Navarra. This project has received funding from the Economic Policy Research Network and the European Research Council under the European Union's Horizon 2020 research and innovation program ERC Starting Grant VALURED (Grant agreement No. 804104) and from the Swiss National Science Foundation project EVAS (Grant agreement No. 10001768), as well as support by the UZH Research Priority Program "Equality of Opportunity."

1 Introduction

What is the relationship between what is good for society and what people *think* is good for society? Does learning about people's ethical views inform which policies ought to be pursued? How?

In this paper, we make progress on these questions by quantifying the extent to which the objective of maximizing social welfare can diverge from the objective of following people's ethical views. We consider a very flexible welfarist framework, in which the "right thing to do" is captured by a social welfare function that is strictly increasing in individual utilities. Crucially, we assume that people's ethical views are an integral part of their utilities. Our main result is that, in this framework, the ranking of allocations must closely mirror people's ethical preferences.

To develop an intuition for our main theoretical result, consider a model in which everyone cares about everyone else; for example, the people of Iowa City have preferences about how consumption should be distributed among the residents of Sacramento. Consequently, the distribution of consumption in Sacramento exerts an externality on the people of Iowa City. This externality is likely to be very small. However, it affects not only the people of Iowa City, but also the people of Dallas, Boston, etc. When there are a lot of other people, this small externality adds up and, for a large enough population, dominates the social welfare calculations.

This result also suggests a welfarist rationale for paternalistic policies. In our model, the people of Iowa City care not only about the distribution of consumption among the people of Sacramento, but also about what each of them consumes. For example, they may want Jayden to eat his vegetables. Jayden's decision to eat sweets instead therefore imposes a small negative externality on the people of Iowa City and the rest of the country. For large populations, the sum of these externalities outweighs any positive effect on Jayden's own utility from eating sweets.

Our results add to the literature on the restrictiveness of the welfarist framework (Sen (1970), Kaplow and Shavell (2001) and Sher (2024)). The closest related paper is Sen (1970), which explores the surprising tension between Pareto and liberalism. When preferences are purely self-regarding, Pareto is fundamentally an anti-paternalistic principle: if Jayden prefers sweets to vegetables and no one else cares what Jayden eats, then, *by Pareto*, we must let Jayden eat his sweets. But, in the presence of other-regarding preferences, the Pareto condition no longer supports the anti-paternalistic approach. Sen's famous result on the impossibility of a Paretian liberal shows that, for certain configurations of preferences, the Pareto principle is incompatible with the liberal ideal that people have rights to certain choices. Here, we show that the problem runs much deeper, since this tension is a robust feature of situations involving large populations of people who have preferences that might conflict with other ethical principles.

This tension relates to the literature on the appropriateness of the Pareto condition in settings where people have other-regarding preferences (see, among others, Fleurbaey (2012), Decerf and der Linden (2016) and Treibich (2019)). Dworkin (1990) and Sher (2020) argue that, in some contexts, it may be better to think of other-regarding preferences as reflecting individuals' moral concern for others, rather than as direct inputs into their own "welfare". Bergstrom (1982) and Milgrom (1993) express a concern that accounting for altruistic preferences leads to double counting of some social costs and benefits. In contrast, Frank (1985) and Kaplow (1995) show that neglecting to take into account other-regarding preferences leads to violations of the standard unanimity condition: that is, everyone agrees that certain transfers are desirable, yet the social ranking rejects them. More recently, Fehr and Charness (2025) have argued, on the basis of empirical and experimental results, that other-regarding preferences are important for the assessment of individual welfare. Støstad and Cowell (2024) show the implications for optimal income taxation when individuals care about income inequality.

The purpose of our paper is not to defend a particular position in this debate. Rather, it is to demonstrate its large stakes. Our results show that even weak other-regarding preferences can considerably narrow down the set of admissible welfarist criteria. As a result, the decision of whether or not we should include ethical preferences as part of individual utilities has drastic implications for how we ought to approach normative analysis. We discuss these implications in our concluding section.

Finally, this paper is related to a growing literature on the economic implications of people's social preferences (see, for example, Dufwenberg, Heidhues, Kirchsteiger, Riedel, and Sobel (2011), Long and Stähler (2012), Aghion, Bénabou, Martin, and Roulet (2023), Dewatripont and Tirole (2024), Kaufmann, Andre, and Kőszegi (2024)), and Nesje (2024). Our question here is normative rather than positive: rather than trying to understand the behavioral implications of social preferences, we study the relationship between people's ethical views and the social welfare function. However, our main result builds on an insight that also plays an important role in this literature (especially Kaufmann et al. (2024)): namely, that tiny externalities that affect many people can have meaningful aggregate implications.

2 Model

Consider a society of $I \in \mathbb{N}$ individuals. Each individual, i, is assigned a bundle $c_i \in \mathbb{R}_+^K$ of private goods, where $K \in \mathbb{N}$ is the finite number of goods. We denote by $c_i^k \geq 0$ the quantity of good k assigned to individual i. Let $\mathbf{c} = (c_1, ..., c_I)$ denote an allocation.

Individual i's preferences over allocations are represented by the utility function

$$U_i(\mathbf{c}) = u(c_i) + w(\mathbf{c}), \tag{1}$$

where u and w are strictly increasing and differentiable. Here, the function u represents the self-regarding component of preferences, whereas w represents people's ethical views about how consumption should be distributed among other people. At this point, we assume that all people have the same u and the same w (that is, the same ethical views). We discuss heterogeneity in Section 5.

Note that the assumption that w is increasing allows only for the case of altruistic preferences.[1] This is somewhat restrictive, as Fehr and Charness (2025) document that a substantial share of people exhibit preferences that violate this condition in various ways. However, the assumption that w is strictly increasing

[1]This is inline with Harsanyi (1982), who argues that only altruistic feelings should count towards one's utility, and not malevolent feelings. See also Fleurbaey (2012).

is not crucial for our analysis. We use it only to guarantee that various marginal rates of substitution are always well-defined. Our results are local and remain valid whenever those marginal rates of substitution exist.

Observation 1. If $u \equiv 0$, then everyone has the same preferences over allocations. In this case, the Pareto ranking is a complete ranking, and the unique Paretian social preference relation is represented by w. Of course, $u \equiv 0$ is an unrealistic benchmark, as it assumes that people are perfectly altruistic (in the sense that people do not exhibit a special preference towards their own consumption). However, this benchmark highlights that, at least in some situations, people's ethical preferences place strong restrictions on the form that Paretian social rankings can take. Even if we think that the ethical views encapsulated in w are misguided, we have no choice but to completely abide by them, unless we are willing to violate the Pareto condition.

It turns out that, even when preferences are highly self-regarding, the other-regarding component of preferences places strong restrictions on the set of plausible welfarist rankings. To derive this result, assume that the social preference relation is represented by a social welfare function of the form

$$W(\mathbf{c}) = V(U_1(\mathbf{c}), ..., U_I(\mathbf{c})) \tag{2}$$

for some strictly increasing and differentiable V. Note that this assumption implies the standard Pareto condition, but is somewhat stronger.[2] The key assumption that we are making is that the only relevant information for ranking allocations in contained in the profile of individual utilities.

[2] We make two additional assumptions: first, that the function V is differentiable, and, second, that it is strictly increasing in all arguments. The differentiability of V is a useful technical assumption, as our results will concern social marginal rates of substitution. When individual preferences are self-regarding, the assumption that V is strictly increasing is implied by the standard Pareto condition: in this case, it is possible to increase the utility of one individual while holding the utilities of other individuals constant; by Pareto, such a variation must result in an increase in social welfare. This logic can be used to show that V is strictly increasing. In the presence of other-regarding preferences, this assumption is stronger, as it may be impossible to vary a single individual's utility without affecting the utilities of other individuals.

Observation 2. Different choices of V correspond to different ways of aggregating individual utilities. When preferences are self-regarding ($w \equiv 0$), this choice has important implications for the social ranking of consumption allocations. For example, in the single-good case, we can specify $V(U_1, ..., U_I) = \sum_i f(U_i)$, where $f(u(c)) = \phi(u^{-1}(u(c))) = \phi(c)$, to obtain a social welfare function that is given by $\sum_i \phi(c_i)$. The concavity of ϕ determines the social aversion to consumption inequality; thus, different ϕs imply very different social rankings. This demonstrates that, when preferences are self-regarding, there is ample room for normative judgments regarding the appropriate degree of inequality aversion, and the choice of V is normatively significant.

Our main result is that, in the presence of other-regarding preferences, the social ranking represented by W cannot deviate too much from people's ethical preferences, which are represented by w. In particular, the choice of V is largely inconsequential.

3 Main result

To state our main result, it is useful to define three concepts. The first is the social marginal rate of substitution. The *social marginal rate of substitution* between c_i^k and $c_{i'}^{k'}$ (at allocation $\mathbf{c}$) is

$$SMRS_{i',k'}^{i,k}(\mathbf{c}) \equiv \frac{\partial W(\mathbf{c})}{\partial c_i^k} \Big/ \frac{\partial W(\mathbf{c})}{\partial c_{i'}^{k'}} .$$

This ratio is informative regarding the social desirability of tradeoffs between individual i's consumption of good k and individual i''s consumption of good k'. For example, consider a policy that increases c_i^k by one small unit but decreases $c_{i'}^{k'}$ by x units. The welfare gain from increasing c_i^k is approximately $\frac{\partial W(\mathbf{c})}{\partial c_i^k}$, while the welfare loss from decreasing $c_{i'}^{k'}$ is approximately $x\frac{\partial W(\mathbf{c})}{\partial c_{i'}^{k'}}$. This policy is socially desirable if and only if the welfare gain exceeds the welfare loss, that is, if and only if

$$\frac{\partial W(\mathbf{c})}{\partial c_i^k} > x\frac{\partial W(\mathbf{c})}{\partial c_{i'}^{k'}} \Longleftrightarrow SMRS_{i',k'}^{i,k}(\mathbf{c}) > x$$

As this condition illustrates, social marginal rates of substitution provide useful guidelines for the normative evaluation of policies that involve marginal consumption tradeoffs.

The second concept is the *impartial marginal rate of substitution* between c_i^k and $c_{i'}^{k'}$ (at allocation $\mathbf{c}$). This marginal rate of substitution is implied by people's other-regarding preferences:

$$IMRS_{i',k'}^{i,k}(\mathbf{c}) \equiv \frac{\partial w(\mathbf{c})}{\partial c_i^k} \Big/ \frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}.$$

Note that the impartial marginal rates of substitution are informative regarding the popular support for policies that involve tradeoffs in c_i^k and $c_{i'}^{k'}$; these correspond to the marginal rates of substitution of "impartial" individuals, that is, all individuals except individuals i and i'.

The third and final concept that we need to define is the *partiality* of individual i with respect to c_i^k (at allocation $\mathbf{c}$). This is defined as

$$p_i^k(\mathbf{c}) = \frac{\partial u(c_i)}{\partial c_i^k} \Big/ \frac{\partial w(\mathbf{c})}{\partial c_i^k}.$$

To motivate this definition, note that, for $i' \neq i$, individual i's marginal rate of substitution between c_i^k and $c_{i'}^k$ is given by

$$\begin{aligned}\frac{\partial U_i(\mathbf{c})}{\partial c_i^k} \Big/ \frac{\partial U_i(\mathbf{c})}{\partial c_{i'}^{k'}} &= \left(\frac{\partial u(c_i)}{\partial c_i^k} + \frac{\partial w(\mathbf{c})}{\partial c_i^k}\right) \Big/ \left(\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}\right) = \\ &= \left(\frac{\partial w(\mathbf{c})}{\partial c_i^k} \Big/ \frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}\right)\left(\frac{\partial u(c_i)}{\partial c_i^k} \Big/ \frac{\partial w(\mathbf{c})}{\partial c_i^k} + 1\right) = \\ &= IMRS_{j,l}^{i,k}(\mathbf{c})\left(p_i^k(\mathbf{c}) + 1\right).\end{aligned}$$

For example, if $p_i^k(\mathbf{c}) = 9$, then individual i values an incremental increase in his own consumption of good k ten times more than an altruistic stranger would value this increase. This means that if most people would be indifferent between increasing individual i's consumption of good k by one unit and increasing individual i''s consumption of good k' by 3 units, then individual i is indifferent between increasing his own consumption of good k by 1 unit and increasing individual i''s consumption of good k' by 30 units. Importantly, $p_i^k(\mathbf{c})$ is a measure of individuals' selfishness, which can be elicited empirically.

The following theorem contains the main result of the paper.

Theorem 1 *Fix an allocation* $\mathbf{c}$*. For each* $n = 1, ..., I-1$*, there exist a set* S *of* n *individuals such that, for any two individuals* $i, i' \in S$ *and any two goods* k, k'*, it holds that*

$$\left(1 + \frac{p_{i'}^{k'}(\mathbf{c})}{I-n}\right)^{-1} \leq \frac{SMRS_{i',k'}^{i,k}(\mathbf{c})}{IMRS_{i',k'}^{i,k}(\mathbf{c})} \leq \left(1 + \frac{p_i^k(\mathbf{c})}{I-n}\right)$$

The proof of this theorem is in the appendix, together with other omitted proofs. To appreciate this result, consider the ratio $SMRS_{i',k'}^{i,k}(\mathbf{c})/IMRS_{i',k'}^{i,k}(\mathbf{c})$. The numerator is the social marginal rate of substitution between c_i^k and $c_{i'}^{k'}$. This quantity represents the ethically-desirable way of trading off incremental increases in c_i^k and $c_{i'}^{k'}$, which depends, of course, on the social welfare function, W. Because the social welfare function is a normative choice, this term has no empirical counterpart. On the other hand, the denominator of this ratio is based on how an impartial individual actually evaluates this tradeoff. This term *is* empirically observable: for example, we can conduct surveys or experiments to learn about how people approach such tradeoffs between strangers.

Our main quantitative insight is that, under mild assumptions, the bounds provided by this theorem can be very tight, and hence we can learn quite a bit about the "correct" marginal rates of substitution from eliciting people's ethical preferences. To see this, note the following:

- The terms $p_i^k(\mathbf{c})$ and $p_{i'}^{k'}(\mathbf{c})$ capture the individuals' willingness to sacrifice their own consumption for the sake of increasing the consumption of a stranger. These terms can be empirically estimated, and bounded by some P – that is, based on empirical evidence, we might be able to conclude that people are willing to sacrifice one dollar of their own income in order to increase a stranger's income by P dollars. Experimental evidence tends to find $P \leq 10$ – that is, most people would prefer an extra \$10 to an equally-well-off stranger over an extra dollar to themselves (see, for example, Charness and Rabin (2002) and Fehr and Charness (2025)). This suggests that

$$p_i^k(\mathbf{c}), p_{i'}^{k'}(\mathbf{c}) \leq 10$$

- When the population is large, it is possible to choose n so that both n and

$(I-n)$ are large. It is useful to have n large so that the bounds apply to many people; it is useful to have $I-n$ large so that the bounds are tight.

To fix ideas, let $I = 10^{10}$, which is roughly the size of the global population. Choose n to equal 99% of the population, so that $I - n = 0.01 * 10^{10} = 10^8$.

- Thus, we have that, for every i and k, $p_i^k(\mathbf{c})/(I-n) \leq 10/10^8 = 10^{-7}$.

Combining, we have that, for 99% of the population, the social marginal rates of substitution must satisfy

$$\frac{1}{1+10^{-7}} \leq \left(1 + \frac{p_{i'}^{k'}(\mathbf{c})}{I-n}\right)^{-1} \leq \frac{SMRS_{i',k'}^{i,k}(\mathbf{c})}{IMRS_{i',k'}^{i,k}(\mathbf{c})} \leq 1 + \frac{p_i^k(\mathbf{c})}{I-n} \leq 1 + 10^{-7}$$

Because 10^{-7} is a tiny number, these bounds are very tight. Even if lab estimates vastly underestimate the degree of partiality, the model still generates tight bounds. For example, as long as $p_i^k(\mathbf{c}) \leq 10^6$ (that is, people prefer a million dollars to a otherwise-identical strangers over one additional dollar to themselves), the social marginal rates of substitution can deviate from the impartial marginal rates of substitution by at most 1% (10^{-2}).

Note that the theorem tells us only that these bounds are relevant for 99% of the population. But it does not tell us which 99%. Without further assumptions, we cannot ensure that these bounds apply to all or any specific subset of individuals. This is because, in this framework, the only ethical requirement is that all individuals matter for social welfare. However, nothing prevents the social welfare function from prioritizing gains to some individuals over gains to others.

3.1 Examples

Our results suggest that different ways of aggregating utilities imply surprisingly similar rankings of consumption distributions. We demonstrate this by comparing the utilitarian social welfare function with a dictatorial social welfare function.

Utilitarianism. Consider the case in which

$$W(\mathbf{c}) = V(U_1(\mathbf{c}), ..., U_I(\mathbf{c})) = \sum_{i=1}^{I} U_I(\mathbf{c})$$

In this case, we have that

$$\frac{\partial W(\mathbf{c})}{\partial c_i^k} = \sum_{j=1}^{I} \frac{\partial U_j(\mathbf{c})}{\partial c_i^k} = \frac{\partial u(\mathbf{c})}{\partial c_i^k} + I\frac{\partial w(\mathbf{c})}{\partial c_i^k}$$

It thus follows that the social marginal rate of substitution between c_i^k and $c_{i'}^{k'}$ is given by

$$\begin{aligned}SMRS_{i',k'}^{i,k}(\mathbf{c}) =& \frac{\frac{\partial u(\mathbf{c})}{\partial c_i^k} + I\frac{\partial w(\mathbf{c})}{\partial c_i^k}}{\frac{\partial u(\mathbf{c})}{\partial c_{i'}^{k'}} + I\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}} = \left(\frac{\frac{\partial w(\mathbf{c})}{\partial c_i^k}}{\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}}\right)\left(\frac{p_i^k(\mathbf{c}) + I}{p_{i'}^{k'}(\mathbf{c}) + I}\right) \\ =& IMRS_{i',k'}^{i,k}(\mathbf{c})\left(\frac{p_i^k(\mathbf{c}) + I}{p_{i'}^{k'}(\mathbf{c}) + I}\right) = IMRS_{i',k'}^{i,k}(\mathbf{c})\left(\frac{\frac{p_i^k(\mathbf{c})}{I} + 1}{\frac{p_{i'}^{k'}(\mathbf{c})}{I} + 1}\right)\end{aligned}$$

In this case, the bounds apply to all individuals. The utilitarian social welfare function ranks consumption allocations approximately according to people's ethical views.

Dictatorial rankings. Consider an alternative in which the aggregation of utilities is locally dictatorial, that is,[3]

$$W(\mathbf{c}) = V(U_1(\mathbf{c}), ..., U_I(\mathbf{c})) = U_1(\mathbf{c}) = u(c_1) + w(\mathbf{c})$$

In this case, for all individuals *except* for individual 1, the social marginal rates of substitution are exactly those implied by w (consistent with our theorem). However, this is not true for individual 1. Here, the function u plays a role as well; in particular, the social welfare function values increases in c_1 much more than an impartial individual would.

The comparison of these example demonstrates that very different utility aggregators must agree, to a large extent, on how to rank consumption distributions. However, they may also differ in how they treat small segments of the population.

[3]It should be noted that such a social welfare function violates our assumption that V is differentiable and strictly increasing in all utilities; however, it is possible to approximate this social welfare function with one that satisfies these properties. In addition, it is worth noting that, locally, this social welfare function is similar to a Rawlsian social welfare function, provided that individual 1 is the worst-off individual.

It is possible to draw more robust conclusions about the social objective if we impose some conditions that guarantee that no individual is too special. Our theorem splits the population into a group for which the social marginal rates of substitution are essentially known (the 99%), and a group of "special" individuals that are needed to keep the bounds tight and might be treated differently (the remaining 1%). However, if we impose further restrictions that guarantee that these special individuals cannot be treated too differently, we can use our theorem to obtain bounds that apply to *all* individuals. In the following section, we illustrate how this might be done by making some assumptions about the social welfare function and the empirically-relevant allocation.

4 Applications

We illustrate the usefulness and implications of Theorem 1 for two types of questions: inequality aversion and paternalism. In both cases, we impose some mild parametric assumptions on the social welfare function, and illustrate how our theorem can be used for obtaining tight bounds on its parameters.

4.1 Inequality aversion

Assume that there is only one good ($K = 1$), and fix some $\mathbf{c}$. Further, assume that, conditional on $\mathbf{c}$, we are willing to restrict attention to social welfare functions for which, for every i,

$$\frac{\partial W(\mathbf{c})}{\partial c_i} = c_i^{-\beta(\mathbf{c})},$$

where $\beta(\mathbf{c})$ is an unknown parameter. This restriction characterizes, locally, an Atkinson social welfare function with inequality parameter $\beta(\mathbf{c})$. In other words, we are assuming that the social marginal rates of substitution between any two individuals should depend only on their relative consumption levels, raised to some power:

$$SMRS_{i'}^{i}(\mathbf{c}) = \left(\frac{c_{i'}}{c_i}\right)^{\beta(\mathbf{c})}$$

The role of this functional-form assumption is to provide a sense in which all individuals must be treated equally. In particular, the social welfare function must consider marginal consumption gains to two individuals with the same consumption level as equally valuable. Moreover, it must apply the same level of inequality aversion across the consumption distribution: for example, if $c_i/c_{i'} = 2$, then the social marginal rate of substitution between i and i' must be $2^{\beta(\mathbf{c})}$, regardless of whether i and i' are at the top of the income distribution or at the bottom of it.

To apply our theorem, assume further that, through surveys or experiments, we find that the impartial marginal rates of substitution are given by

$$IMRS^i_{i'}(\mathbf{c}) = \left(\frac{c_{i'}}{c_i}\right)^{b(\mathbf{c})},$$

where $b(\mathbf{c})$ is a the degree of people's inequality aversion.

Our results tell us that, for 99% of the population, it must hold that

$$\frac{1}{1+10^{-7}} \leq \frac{SMRS^i_{i'}(\mathbf{c})}{IMRS^i_{i'}(\mathbf{c})} = \left(\frac{c_{i'}}{c_i}\right)^{\beta(\mathbf{c})-b(\mathbf{c})} \leq 1+10^{-7}$$

Unless consumption is very equally distributed (so that $c_{i'}/c_i \approx 1$ for most individuals), this condition places tight restrictions on $\beta - b$. To illustrate, assume that in any group containing 99% of the population, there are two individuals, i and i', such that the ratio of their consumptions is at least $x > 1$. So, we must have that

$$\frac{1}{1+10^{-7}} \leq \left(\frac{c_i}{c_{i'}}\right)^{\beta(\mathbf{c})-b(\mathbf{c})} \leq 1+10^{-7}$$

Taking logs, we have that

$$-\ln(1+10^{-7}) \leq (\beta(\mathbf{c}) - b(\mathbf{c}))\ln\left(\frac{c_i}{c_{i'}}\right) \leq \ln(1+10^{-7})$$

Given that $\left(\frac{c_i}{c_{i'}}\right) \geq x > 1$, we obtain

$$-\frac{\ln(1+10^{-7})}{\ln(x)} \leq -\frac{\ln(1+10^{-7})}{\ln\left(\frac{c_i}{c_{i'}}\right)} \leq (\beta(\mathbf{c}) - b(\mathbf{c})) \leq \frac{\ln(1+10^{-7})}{\ln\left(\frac{c_i}{c_{i'}}\right)} \leq \frac{\ln(1+10^{-7})}{\ln(x)}$$

Note that a higher x corresponds to more consumption inequality, and tighter bounds on the deviation $(\beta - b)$. However, even very modest amounts of inequality imply tight bounds. For example, if $x = 1.1$ (that is, it is impossible to find a group of 99% of the population in which everyone earns within 10% of each other), we have

$$-10^{-6} \approx -\frac{\ln(1+10^{-7})}{\ln(1.1)} \leq (\beta(\mathbf{c}) - b(\mathbf{c})) \leq \frac{\ln(1+10^{-7})}{\ln(1.1)} \approx 10^{-6}$$

Implying that the normative parameter, $\beta(\mathbf{c})$, cannot deviate much from the empirically-estimable one, $b(\mathbf{c})$.[4]

Relationship to Harsanyi. The standard approach in applied welfare analysis is to relate social inequality aversion to individual risk aversion (see Harsanyi (1953) and Harsanyi (1955)). The idea is that individuals face a "birth lottery" that determines their consumption level after birth. Higher inequality means that the lottery is more risky, because people do not know whether they will be born in the top of the income distribution or in the bottom it. Behind the veil of ignorance, the level of risk aversion determines individuals' attitudes towards these birth lotteries. The Pareto condition can then be used to fully characterize a social ranking that reflects these attitudes towards risk (see also Eden (2020)).

Our results here suggest that, even within Harsanyi's framework, the social welfare function closely mirrors people's attitudes towards inequality among strangers, while attitudes towards risk play a much more minor role. To illustrate, it is useful to map our framework into Harsanyi's. For this purpose, we can think of the K different goods as representing consumption in K different states, and assume that individuals are expected utility maximizers, so that

$$u_i(c_i) = \sum_{k=1}^{K} p_k u^{vNM}(c_i^k) \text{ and } w(\mathbf{c}) = \sum_{k=1}^{K} p_k w^{vNM}(c_1^k, ..., c_I^k),$$

[4]It may be useful to contrast this result with Støstad and Cowell (2024). There, it is assumed that other-regarding preferences are concerned exclusively with inequality, rather than with the welfare of individuals in society. Unlike here, the concern for inequality does not fully determine the way that the planner evaluates allocations, while still having important implications for policy design. The main difference is that, in Støstad and Cowell (2024), the degree of partiality of an individual—that is, the relative importance of self vs strangers—goes to infinity with the number of individuals in society, while here it is assumed to be finite.

where $u^{vNM} + w^{vNM}$ is the individual's von Neumann and Morgenstern utility function, and p_k is the probability of state k. Harsanyi's social welfare function is given by the expected utility of the birth lottery, which is

$$\frac{1}{I}\sum_{i=1}^{I} U_i(\mathbf{c}) = \frac{1}{I}\sum_{i=1}^{I}\left(u^{vNM}(c_i^k) + w^{vNM}(c_1^k, ..., c_I^k)\right) =$$

$$\left(\frac{1}{I}\sum_{i=1}^{I} u^{vNM}(c_i^k)\right) + w^{vNM}(c_1^k, ..., c_I^k),$$

where c_i^k is the individual's consumption level in the realized state, k. We thus have that Harsanyi's social welfare function has two components: the first corresponding to individuals' attitudes towards their own consumption risk, and the second corresponding to their ethical views. The literature has mostly focused on the first component, ignoring the second one. However, our quantitative analysis here suggests that the second component may end up dominating the social welfare calculations.

4.2 Paternalism

Consider the case in which there are at least two goods ($K \geq 2$). Goods 1 and 2 are considered by everyone to be perfect substitutes: for example, we can think of low-fat milk and whole milk, or sugared soda and sparkling water. Further, assume that everyone thinks that everyone else should approach the tradeoffs between these two goods in the same way; that is, there exists some $\gamma > 0$ such that, for every i,

$$IMRS_{i,1}^{i,2}(\mathbf{c}) = \gamma$$

Furthermore, assume that we are willing to restrict attention to social welfare criteria that, conditional on $\mathbf{c}$, (a) treat goods 1 and 2 as perfect substitutes (locally), and (b) evaluate local tradeoffs between these two goods in the same way for all individuals. In particular, assume that there exists a parameter, γ^*, so that

$$SMRS_{i,1}^{i,2}(\mathbf{c}) = \gamma^*$$

Here, the assumption that the *same* γ^* applies to all individuals will allow us to say that bounds on γ^*/γ that apply to 99% of the population should also apply to the remaining 1%.

In particular, our theorem implies that, for 99% of the population (that is, $i \in S$),

$$\frac{1}{1+10^{-7}} \leq \frac{SMRS_{i,1}^{i,2}(\mathbf{c})}{IMRS_{i,1}^{i,2}(\mathbf{c})} = \frac{\gamma^*}{\gamma} \leq 1 + 10^{-7}$$

These bounds establish that γ^* cannot deviate from γ by more than 10^{-7}. As, given our assumptions, the social and impartial marginal rates of substitution between goods 1 and 2 are the same for *all* individuals, we can conclude that these bounds apply to all individuals, and not just those in S.

It is worth highlighting that, in this example, the social marginal rates of substitution are completely insensitive to individuals' preferences over the composition of their own consumption bundles. For example, individuals may prefer to drink sugared soda over sparkling water (unless the latter is prohibitively expensive relative to the former); but, as long as everyone thinks that everyone else should be drinking the water instead, the social welfare function must adopt this paternalistic view.

5 Heterogeneous ethical views

Before concluding our analysis, it is worth clarifying the implications of introducing heterogeneous other-regarding preferences. Not surprisingly, doing so allows for more flexibility regarding the social welfare criterion. Most trivially, this flexibility emerges because the social welfare function can weigh individuals differently depending on their ethical views.

Loosely, our main result can be generalized as, "With heterogeneous ethical views, social welfare cannot differ too much from the ethical views of individuals." More precisely, the social marginal rate of substitution must be "close" to some linear combination of the impartial marginal rate of substitutions of the individuals.

To formally state this result, consider an extension in which individual preferences are heterogeneous. This means that individual preferences are now represented by the utility functions

$$U_i(\mathbf{c}) = u_i(c_i) + w_i(\mathbf{c}), \tag{3}$$

where u_i and w_i are strictly increasing and differentiable. The main extension is that, here, we allow w to be differ across individuals. However, once we do that, it doesn't make a big difference if we allow u to be heterogeneous across individuals as well.

Let $\vec{\alpha} = (\alpha_1, ..., \alpha_I)$ be a vector of positive weights (that sum up to 1). Define the weighted (mediant of the) impartial marginal rates of substitutions (given $\vec{\alpha}$) as

$$\overline{IMRS}_{i',k'}^{i,k}(\mathbf{c},\vec{\alpha}) \equiv \frac{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k}}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}}} = \frac{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}} \left(\left(\frac{\partial w_j(\mathbf{c})}{\partial c_i^k} \right) / \left(\frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}} \right) \right)}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}}}$$

Note that $\overline{IMRS}_{i',k'}^{i,k}$ is bounded by the smallest and largest impartial marginal rate of substitution of individuals. Similarly, define the weighted (mediant) partiality of individual i (conditional on $\vec{\alpha}$) as

$$\overline{p}_i^k(\mathbf{c},\vec{\alpha}) = \frac{\partial u_i(c_i)}{\partial c_i^k} \Bigg/ \left(\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k} \right).$$

This corresponds to a notion of partiality that is relative to some weighted average of the population: that is, a person's extra concern for their own consumption relative to a weighted average of others' concerns for them.

Theorem 1 extends when replacing the impartial marginal rate of substitution and the partiality parameters with their weighted counterparts.

Theorem 2 *Fix some* $\mathbf{c}$*. For each* $n = 1, ..., I-1$*, there exist a set* S *of* n *individuals and weights* $\vec{\alpha} = (\alpha_1, ..., \alpha_I)$ *such that, for any two individuals* $i, i' \in S$ *and any two goods* k, k'*, it holds that*

$$\left(1 + \frac{\overline{p}_{i'}^{k'}(\mathbf{c},\vec{\alpha})}{I-n}\right)^{-1} \leq \frac{SMRS_{i',k'}^{i,k}(\mathbf{c})}{\overline{IMRS}_{i',k'}^{i,k}(\mathbf{c},\vec{\alpha})} \leq \left(1 + \frac{\overline{p}_i^k(\mathbf{c},\vec{\alpha})}{I-n}\right).$$

This result establishes that, even with heterogeneity, the social marginal rates of substitution are constrained by people's other-regarding preferences: social preferences cannot deviate too much from a compromise between the heterogeneous ethical views of individuals. Unsurprisingly, Pareto efficiency is insufficient for identifying how the heterogeneous ethical views of individuals ought to be aggregated.

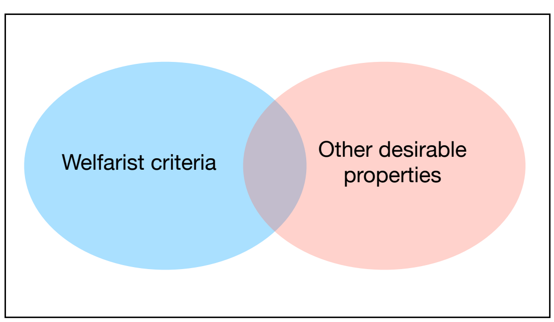
However, it is worth noting that if, locally, the impartial marginal rates of substitution are the same across individuals, then the bounds of Theorem 1 apply. In particular, local consensus in ethical views implies tight bounds on the social marginal rates of substitution.

6 Conclusion

In this concluding section, we discuss the broader implications of our result for the study of normative economics.

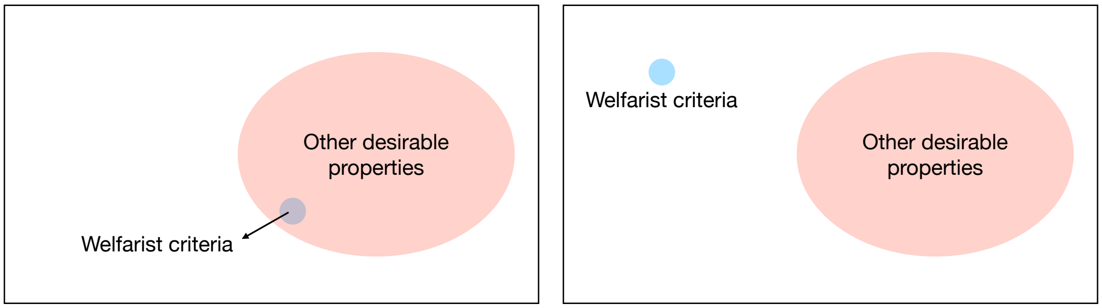
When individual welfare depends only on one's own situation (Figure 1), the key question is how to aggregate the utilities of individuals. There are many social orderings that are increasing in individual utilities, and the problem is to choose the right one. While some ethical principles turn out to be incompatible with this welfarist framework, many others can be expressed through the way in which we choose to aggregate utilities. For example, a concern for fairness can be expressed with the requirement that the social aggregation of individual utilities should be symmetric and inequality averse. Formulating such requirements is helpful for narrowing down the set of admissible criteria.

Figure 1: Framework with self-regarding preferences



In contrast, when individuals' utilities depend on the entire allocation (Figure 2), the problem faced by the normative economist is of an entirely different

Figure 2: Framework with other-regarding preferences



(a) Optimistic scenario (b) Pessimistic scenario

nature. Our results show that the set of welfarist criteria—or at least those that are representable by a social welfare function that is strictly increasing in individuals' utilities—vanishes. Consequently, the way we aggregate individual utilities makes little difference in terms of how we rank allocations. This means that there is limited scope for expressing other ethical principles through the choice of aggregation.

In this case, we find ourselves in one of two situations. In the optimistic scenario (Figure 2a), people's ethical views are already consistent with all of the desirable properties that we would like to impose on our normative criterion. For example, it could be that people's preferences are paternalistic only in cases in which paternalism is justified; that their attitudes towards inequality are within the acceptable range; etc. In this case, Pareto efficiency provides us with a reason to select, among all such criteria, one that closely aligns with people's ethical views. This observation turns the normative question into an empirical one: the social ordering can be inferred based on how people assess normative tradeoffs.

However, we might also find ourselves in a more pessimistic scenario, as in Figure 2b. In this case, people's ethical views are, in some way, in conflict with social desiderata. For example, perhaps people are racist, even though racism is clearly wrong. In this case, we need to think about how to weaken Pareto efficiency or balance it with other desirable properties.

References

Aghion, P., Bénabou, R., Martin, R., Roulet, A. (2023). Environmental preferences and technological choices: Is market competition clean or dirty? *American Economic Review: Insights*, *5*(1), 1–19.

Bergstrom, T. (1982). When is a man's life worth more than his human capital? In M. Jones-Lee (Ed.), *The value of life and safety.* Amsterdam: North-Holland.

Charness, G., Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*(3), 817–869.

Decerf, B., der Linden, M. V. (2016). Fair social orderings with other-regarding preferences. *Social Choice and Welfare*, *46*, 655–694.

Dewatripont, M., Tirole, J. (2024). The morality of markets. *Journal of Political Economy*, *132*(8), 2655–2694.

Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., Sobel, J. (2011). Other-regarding preferences in general equilibrium. *The Review of Economic Studies*, *78*(2), 613–639.

Dworkin, R. (1990). The 'double counting' objection. In J. Glover (Ed.), *Utilitarianism and its critics* (p. 103-106). MacMillan.

Eden, M. (2020, December). Welfare analysis with heterogeneous risk preferences. *Journal of Political Economy*, *128*(12), 4574–4613.

Fehr, E., Charness, G. (2025). Social preferences: fundamental characteristics and economic consequences. *Journal of Economic Literature*, *63*(2), 440–514.

Fleurbaey, M. (2012). The importance of what people care about. *Politics, Philosophy and Economics*, *11*, 415–447.

Frank, R. H. (1985). The demand for unobservable and other nonpositional goods. *The American Economic Review*, *75*(1), 101–116.

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, *61*(5), 434–435.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*(4), 309–321.

Harsanyi, J. C. (1982). Morality and the theory of rational behaviour. In S. AK

W. B (Eds.), *Utilitarianism and beyond.* Cambridge: Maison des Sciences de l'Homme and Cambridge University Press.

Kaplow, L. (1995). A note on subsidizing gifts. *Journal of Public Economics*, *58*(3), 469–477.

Kaplow, L., Shavell, S. (2001). Any non-welfarist method of policy assessment violates the pareto principle. *Journal of Political Economy*, *109*(2), 281–286.

Kaufmann, M., Andre, P., Kőszegi, B. (2024). Understanding markets with socially responsible consumers. *The Quarterly Journal of Economics*, *139*(3), 1989–2035.

Long, N. V., Stähler, F. (2012). Should the good and the selfish be taxed differently? *The Scandinavian Journal of Economics*, *114*(3), 932–948.

Milgrom, P. (1993). Is sympathy an economic value? philosophy, economics, and the contingent valuation method. *Contingent valuation: A critical Assessment*, 417–441.

Nesje, F. (2024). Cross-dynastic intergenerational altruism. *Journal of Environmental Economics and Management*, *128*, 103059.

Sen, A. (1970). The impossibility of a paretian liberal. *Journal of Political Economy*, *78*(1), 152–157.

Sher, I. (2020). How perspective-based aggregation undermines the pareto principle. *Politics, Philosophy & Economics*, *19*(2), 182–205.

Sher, I. (2024). Generalized social marginal welfare weights imply inconsistent comparisons of tax policies. *American Economic Review*, *114*, 3551–3577.

Støstad, M. N., Cowell, F. (2024). Inequality as an externality: Consequences for tax design. *Journal of Public Economics*, *235*, 105139. doi: https://doi.org/10.1016/j.jpubeco.2024.105139

Treibich, R. (2019). Welfare egalitarianism with other-regarding preferences. *Social Choice and Welfare*, *52*, 1–28.

A Proof of Theorem 1

$$\frac{\partial W(\mathbf{c})}{\partial c_i^k} = \sum_{j=1}^{I} \frac{\partial V(U_1(\mathbf{c}), ..., U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})} \frac{\partial U_j(\mathbf{c})}{\partial c_i^k} =$$

$$\left(\sum_{j=1}^{I}\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}\frac{\partial w(\mathbf{c})}{\partial c_i^k}\right)+\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_i(\mathbf{c})}\frac{\partial u(c_i)}{\partial c_i^k}=$$

$$\frac{\partial w(\mathbf{c})}{\partial c_i^k}\left(\sum_{j=1}^{I}\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}\right)+\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_i(\mathbf{c})}\frac{\partial u(c_i)}{\partial c_i^k}$$

It follows that the social marginal rates of substitution are given by

$$\frac{\frac{\partial W(\mathbf{c})}{\partial c_i^k}}{\frac{\partial W(\mathbf{c})}{\partial c_{i'}^{k'}}}=\frac{\frac{\partial w(\mathbf{c})}{\partial c_i^k}\left(\sum_{j=1}^{I}\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}\right)+\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_i(\mathbf{c})}\frac{\partial u(c_i)}{\partial c_i^k}}{\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}\left(\sum_{j=1}^{I}\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}\right)+\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_{i'}(\mathbf{c})}\frac{\partial u(c_{i'})}{\partial c_{i'}^{k'}}}=$$

$$\left(\frac{\frac{\partial w(\mathbf{c})}{\partial c_i^k}}{\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}}\right)\left(\frac{1+\left(\frac{\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_i(\mathbf{c})}}{\sum_{j=1}^{I}\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}}\right)\left(\frac{\frac{\partial u(c_i)}{\partial c_i^k}}{\frac{\partial w(\mathbf{c})}{\partial c_i^k}}\right)}{1+\left(\frac{\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_i(\mathbf{c})}}{\sum_{j=1}^{I}\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}}\right)\left(\frac{\frac{\partial u(c_{i'})}{\partial c_{i'}^{k'}}}{\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}}\right)}\right)=$$

$$\left(\frac{\frac{\partial w(\mathbf{c})}{\partial c_i^k}}{\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}}\right)\left(\frac{1+\frac{\alpha_i}{\sum_{j=1}^{I}\alpha_j}p_i^k(\mathbf{c})}{1+\frac{\alpha_{i'}}{\sum_{j=1}^{I}\alpha_j}p_{i'}^{k'}(\mathbf{c})}\right)$$

where

$$\alpha_j=\frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}$$

Hence,

$$\frac{SMRS_{i',k'}^{i,k}}{IMRS_{i',k'}^{i,k}}=\frac{\left(\frac{\frac{\partial w(\mathbf{c})}{\partial c_i^k}}{\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}}\right)\left(\frac{1+\frac{\alpha_i}{\sum_{j=1}^{I}\alpha_j}p_i^k(\mathbf{c})}{1+\frac{\alpha_{i'}}{\sum_{j=1}^{I}\alpha_j}p_{i'}^{k'}(\mathbf{c})}\right)}{\left(\frac{\frac{\partial w(\mathbf{c})}{\partial c_i^k}}{\frac{\partial w(\mathbf{c})}{\partial c_{i'}^{k'}}}\right)}=\frac{1+\frac{\alpha_i}{\sum_{j=1}^{I}\alpha_j}p_i^k(\mathbf{c})}{1+\frac{\alpha_{i'}}{\sum_{j=1}^{I}\alpha_j}p_{i'}^{k'}(\mathbf{c})} \tag{4}$$

Construct S as the set of n individuals with the lowest α_j's. For any i that is in this set, it holds that

$$\frac{\alpha_i}{\sum_{j=1}^{I}\alpha_j}<\frac{\alpha_i}{\sum_{j=\in\{1,...,I\}\setminus S}\alpha_j}\leq\frac{\alpha_i}{(I-n)\alpha_i}=\frac{1}{I-n}$$

Thus, for $i,i'\in S$, equation (4) implies that

$$\frac{1}{1+\frac{p_{i'}^{k'}(\mathbf{c})}{I-n}}<\frac{1}{1+\frac{\alpha_{i'}}{\sum_{j=1}^{I}\alpha_j}p_{i'}^{k'}(\mathbf{c})}<\frac{SMRS_{i',k'}^{i,k}}{IMRS_{i',k'}^{i,k}}<1+\frac{\alpha_i}{\sum_{j=1}^{I}\alpha_j}p_i^k(\mathbf{c})<1+\frac{p_i^k(\mathbf{c})}{I-n}$$

B Proof of Theorem 2

$$\frac{\partial W(\mathbf{c})}{\partial c_i^k} = \sum_{j=1}^{I} \frac{\partial V(U_1(\mathbf{c}), ..., U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})} \frac{\partial U_j(\mathbf{c})}{\partial c_i^k} =$$

$$\left(\sum_{j=1}^{I} \frac{\partial V(U_1(\mathbf{c}), ..., U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})} \frac{\partial w_j(\mathbf{c})}{\partial c_i^k} \right) + \frac{\partial V(U_1(\mathbf{c}), ..., U_I(\mathbf{c}))}{\partial U_i(\mathbf{c})} \frac{\partial u_i(c_i)}{\partial c_i^k}$$

let $\alpha \equiv \sum_{j=1}^{I} \frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})}$ and $\alpha_j \equiv \frac{\partial V(U_1(\mathbf{c}),...,U_I(\mathbf{c}))}{\partial U_j(\mathbf{c})} / \alpha$. Then

$$\frac{\partial W(\mathbf{c})}{\partial c_i^k} = \alpha \left(\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k} + \alpha_i \frac{\partial u_i(c_i)}{\partial c_i^k} \right).$$

It follows that the social marginal rates of substitution are given by

$$\frac{\frac{\partial W(\mathbf{c})}{\partial c_i^k}}{\frac{\partial W(\mathbf{c})}{\partial c_{i'}^{k'}}} = \frac{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k} + \alpha_i \frac{\partial u_i(c_i)}{\partial c_i^k}}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}} + \alpha_{i'} \frac{\partial u_{i'}(c_{i'})}{\partial c_{i'}^{k'}}} =$$

$$\left(\frac{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k}}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}}} \right) \left(\frac{1 + \alpha_i \frac{\sum_{j=1}^{I} \alpha_j \frac{\partial u_i(c_i)}{\partial c_i^k}}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k}}}{1 + \alpha_{i'} \frac{\frac{\partial u_{i'}(c_{i'})}{\partial c_{i'}^{k'}}}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}}}} \right).$$

Let the weighted mediant of the impartial marginal rates of substitutions be

$$\overline{IMRS}_{i',k'}^{i,k} \equiv \frac{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k}}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_{i'}^{k'}}}$$

and let the weighted mediant of the partiality of i be

$$\overline{p}_i^k(\mathbf{c}) \equiv \frac{\sum_{j=1}^{I} \alpha_j \frac{\partial u_i(c_i)}{\partial c_i^k}}{\sum_{j=1}^{I} \alpha_j \frac{\partial w_j(\mathbf{c})}{\partial c_i^k}}.$$

By construction, $\overline{IMRS}$ is between the smallest and largest $IMRS$ of individuals. Hence,

$$\frac{SMRS_{i',k'}^{i,k}}{\overline{IMRS}_{i',k'}^{i,k}} = \frac{1 + \alpha_i \overline{p}_i^k(\mathbf{c})}{1 + \alpha_{i'} \overline{p}_{i'}^{k'}(\mathbf{c})} \tag{5}$$

Construct S as the set of n individuals with the lowest α_j's. For any i that is in this set, it holds that

$$\alpha_i \leq \frac{1}{I-n}.$$

Thus, for $i, i' \in S$, equation (5) implies that:

$$\frac{1}{1+\frac{\overline{p}_{i'}^{k'}(\mathbf{c})}{I-n}} \leq \frac{1}{1+\alpha_{i'}\overline{p}_{i'}^{k'}(\mathbf{c})} \leq \frac{SMRS_{i',k'}^{i,k}}{\overline{IMRS}_{i',k'}^{i,k}} \leq 1+\alpha_i\overline{p}_i^k(\mathbf{c}) \leq 1+\frac{\overline{p}_i^k(\mathbf{c})}{I-n}.$$