The Erasmus Journal for Philosophy and Economics (EJPE) is a peer-reviewed bi-annual academic journal supported by the Erasmus Institute for Philosophy and Economics, Erasmus School of Philosophy, Erasmus University Rotterdam. EJPE publishes research on the methodology of economics, history of economic thought, ethics and economics, and the conceptual analysis of inter-disciplinary work relating economics to other fields. EJPE is an open-access journal. For additional information, see our website: <http://ejpe.org>.

EDITORS
Måns Abrahamson
Annalisa Costella
Savriél Dillingh
Akshath Jitendranath
Erica Celine Yu

ADVISORY BOARD

EDITORIAL ASSISTANTS
James Grayot, Bronagh Dunne

ISSN: 1876-9098
## TABLE OF CONTENTS

### ARTICLES

- **Measuring Freedom:** Towards a Solution to John Rawls’ Indexing Problem  
  *Thomas Ferretti*  
  [pp. 1–31]

- **The Case of Stated Preferences and Social Well-Being Indices**  
  *Shiri Cohen Kaminitz and Iddan Sonsino*  
  [pp. 32–55]

- **“The Hardest of All the Problems”: Hochman, Rodgers, and Buchanan on Pareto Optimal Redistribution**  
  *Daniel Kuehn*  
  [pp. 56–82]

- **Why ‘Indirect Discrimination’ Is a Useful Legal but Not a Useful Moral Concept**  
  *Kasper Lippert-Rasmussen*  
  [pp. 83–107]

### INTERVIEW

- **The Philosophy and Economics of Measuring Discrimination and Inequality:** A Roundtable Discussion  
  *Kasper Lippert-Rasmussen, Xavier Ramos, and Dirk Van de Gaer*  
  [pp. 108–123]

### BOOK REVIEWS

- **Joshua L. Cherniss’s *Liberalism in Dark Times: The Liberal Ethos in the Twentieth Century***  
  *Jon Murphy*  
  [pp. 124–129]

- **Randall G. Holcombe’s *Coordination, Cooperation, and Control: The Evolution of Economic and Political Power***  
  *Vaios Koliofotis*  
  [pp. 130–137]
Avia Pasternak’s *Responsible Citizens, Irresponsible States: Should Citizens Pay for Their States’ Wrongdoings?*
*SOLMU ANTTLA* [pp. 138–144]

Keith Tribe’s *Constructing Economic Science: The Invention of a Discipline 1850–1950*
*ERWIN DEKKER* [pp. 145–152]

**PHD THESIS SUMMARIES**

Thresholds and Limits in Theories of Distributive Justice
*DICK TIMMER* [pp. 153–156]

Desert, Luck, and Justice
*HUUB BROUWER* [pp. 157–161]

Choice Architecture: A Message and Environment Perspective
*LUCA CONGIU* [pp. 162–166]

Unfair Inequality: From Measurement to Causal Drivers
*PAUL HUFNE* [pp. 167–171]

Issues on the Measurement of Opportunity Inequality
*HUGO DEL VALLE-INALÁN CRUCES* [pp. 172–178]

Three Economic Extensions of John Rawls’s Social Contract Theory: European Union, Tax Compliance, and Climate Change
*KLAUDIJO KLASER* [pp. 179–183]

*VALENTINA ERASMO* [pp. 184–193]
Measuring Freedom: Towards a Solution to John Rawls’ Indexing Problem

THOMAS FERRETTI
London School of Economics

Abstract: Suppose a principle of distributive justice states that social institutions should maximize the freedom of the least well-off. Understanding how to do so would be easier if freedom only depended on one good, like income. If it depends instead on a composite index of social primary goods, a question arises: Which combination of social primary goods can maximize the freedom of the least well-off? This is John Rawls’ indexing problem. Solving it requires addressing two related problems. The first consists in evaluating, in theory, under which conditions it is acceptable to substitute goods, that is, their substitution rates. The second consists in evaluating which acceptable substitutions are feasible in practice. This article proposes a framework to think clearly about this indexing problem within a Rawlsian, resourcist conception of distributive justice. I conclude by discussing a path towards solving the indexing problem. While further empirical exploration is needed, plausible assumptions about social regimes suggest that maximizing the freedom of the least well-off is likely to require giving them access to a social position with a balanced combination of social primary goods.

Keywords: freedom, indexing problem, John Rawls, social primary goods, substitution

JEL Classification: A13, D3, D63, I32, J81, P51

I. INTRODUCTION

According to John Rawls’ liberal egalitarian theory of justice, the problem of distributive justice consists in finding how to regulate social institutions to realize a fair distribution of the means of freedom produced by

AUTHOR’S NOTE: This paper benefited from discussions with audiences at the universities of Princeton, Aarhus, Louvain, and the London School of Economics. I am particularly grateful to David V. Axelsen, Pier-André Bouchard St-Amant, Koen Decancq, Marc Fleurbaey, Axel Gossery, Philippe Van Parijs, Alex Voorhoeve, Danielle Zwarthoed, and my former colleagues at the Hoover Chair of Economic and Social Ethics for their helpful comments at various stages of research and writing. I thank Agnes Ofner for helping with the graphs. Between 2016–2018, early research required for this article was funded by the Fonds de Recherche du Québec Société et Culture (Grant number: 197851).
social cooperation. In other words, the metric of justice used to compare individual positions in society is the freedom of individuals in such positions to advance their ends and the means of such freedom are what Rawls calls social primary goods, that is, the basic liberties and social resources allowing free and equal citizens to advance their ends (Rawls 2001, 50–52, 57–61). For Rawls, some means of freedom such as basic liberties must be equal while others, such as income and wealth, can vary in their distribution. In the latter case, inequalities are acceptable if and only if they satisfy two conditions: they maximize the benefits of the least well-off and are attached to positions open to all under fair equality of opportunity (Rawls 2001, 42–43). So, once institutional rights and liberties are secured equally and fair equality of opportunity is guaranteed, institutions of the basic structure must distribute the other means of freedom to the greatest benefit of the least well-off. But understanding how to do so would be easier if freedom only depended on one good, like income. If it depends instead on a composite index of social primary goods, a question arises: Which combination of social primary goods can maximize the freedom of the least well-off? This is Rawls’ indexing problem.

This problem is formulated in slightly different ways in the literature. For instance, John Rawls adopts the standpoint of a representative individual in the least well-off position and asks, “Which combination of primary social goods it would be rational for [the least well-off] to prefer?” (Rawls 1971, 94), while Richard Arneson asks more generally: “How to aggregate a person’s holdings of the various primary goods into a measure of her overall share of them?” (Arneson 1990, 445–446). Suppose that the freedom of the least well-off is affected by two goods: income and power. Does more income compensate for less power? If so, all substitutions might not preserve the value of the bundle: What is the substitution rate? And does the answer depend on each person’s preferences?

To avoid this indexing problem, authors like John Tomasi and Samuel Freeman (at least in his later work, see Freeman 2013) propose to only consider or give priority to one good, either income or economic powers. Arneson argues instead that any metric must respect reasonable

---

1 In what follows, I use ‘resources’ and ‘social primary goods’ interchangeably.
2 I occasionally simplify Rawls’ term ‘income and wealth’ by referring only to ‘income’.
3 Tomasi insists that only income matters in increasing the worth of freedoms for individuals (Tomasi 2012, 190–191), while Freeman insists that the fair opportunity to exercise our economic powers should have priority over the difference principle and increasing the income and wealth of the least well-off (Freeman 2013, 31–32). By ‘economic powers’, he has in mind powers attached to positions of authority in the economy, such as organizing work within a firm, making impactful management decisions, and so on.
pluralism because people may disagree about the value of different goods. For him, “there really is no alternative to subjectivist standards of distributive justice—standards that let the valuation of resource shares depend on the evaluation that each individual herself gives to her share” (Arneson 1990, 446; see also Fleurbaey 2007, 633–636; 2008, 9). But subjective, welfarist metrics raise problems since people with different subjective evaluations of the least well-off’s share may disagree about what fair institutional arrangements are.

The main contribution of this article is to propose a framework to think clearly about the indexing problem within a Rawlsian, resourcist conception of distributive justice. Solving it requires addressing two related problems. The first consists in evaluating, in theory, under which conditions it is acceptable to substitute social primary goods—that is, their substitution rates. I argue that social primary goods have diminishing marginal returns and, therefore, diminishing marginal substitution rates: other things equal, the less of a good there is in the bundle, the higher its substitution rate gets. The second problem consists in evaluating which acceptable substitutions are feasible in practice. I argue that, because of scarcity constraints, some substitutions may be acceptable in theory but not feasible in practice, and the higher substitution rates get, the more difficult it is to afford substitutions.

While I mainly aim at proposing this framework, I conclude by discussing a path towards solving Rawls’ indexing problem. Further empirical exploration is needed for a full demonstration, but plausible assumptions about social regimes and the feasibility of substitutions suggest that maximizing the freedom of the least well-off is likely to require giving them access to a social position with a balanced combination of all social primary goods. This could open new lines of argument against authors proposing to give priority to one good, as Tomasi and Freeman do, for instance to reduce distributive justice to the distribution of income to the least well-off. Indeed, Tomasi justifies a kind of free market system partly on the capacity of this regime to improve the income of the least well-off, but other regimes may give the least well-off access to a more balanced bundle of all social primary goods that maximizes their overall freedom.

The article proceeds as follows. In section II, I review three important features of Rawls’ resourcist metric of justice. In section III, I start discussing the first problem and explain how to understand the substitution of goods and their substitution rates. In section IV, I discuss one way to evaluate the substitution rates of social primary goods. In section V, I turn
to the second problem, and I investigate how scarcity constraints affect the feasibility of substitutions. In section VI, I discuss how, given plausible assumptions about social regimes and the feasibility of substitutions, the indexing problem may have a solution. In section VII, I conclude.

II. RAWLS’ RESOURCIST METRIC OF JUSTICE

I set aside issues regarding the protection of basic liberties. Conditions for their adequate protection are defined by Rawls’ first principle: “Each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties” (Rawls 2001, 42). For Rawls, issues regarding basic liberties should be treated independently:

The fundamental liberties are always equal [...]; one does not need to balance these liberties and rights against other values. The primary social goods that vary in their distribution are the powers and prerogatives of authority, and income and wealth. (Rawls 1971, 93)

In this article, I focus on the social primary goods that vary in their distribution. Regarding these goods, Rawls’ second principle claims that:

Social and economic inequalities are to satisfy two conditions: first, they are to be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least-advantaged members of society (the difference principle). (Rawl 2001, 42–43; see also Rawls 1971, 83)

For the sake of simplicity, I adopt one specific interpretation of Rawls’ resourcist metric of distributive justice as a starting point to illustrate the indexing problem. My primary goal is not to get into exegetical arguments over the right interpretation of the Rawlsian metric, and my account may depart from usual accounts on some occasions. But assuming this specific interpretation of a resourcist metric, I focus on investigating what it means to maximize the value of the bundle of social primary goods of the least well-off, with the hope that this can inform any resourcist conception. Moreover, my account is vulnerable to external objections against resourcist metrics in general, but I will not discuss these in detail. Now, with these caveats in mind, I need to introduce three important features of a plausible resourcist conception of justice to provide sufficient context for my discussion of the indexing problem.

Rawls (1971, 93) also sets aside basic liberties when discussing the indexing problem.

First, measuring how much freedom, that is, how many opportunities, individuals have to advance their ends requires weighing a variety of means of freedom—that is, the basic liberties and social resources allowing free and equal citizens to advance their ends (Rawls 2001, 50–52, 57–61). Indeed, freedom can be defined as the capacity of an agent to realize a chosen action without constraint (MacCallum 1967, 314). Under this conception, freedom is a matter of degree: it depends on the capacity of the agent, which depends on their external resources, such as income and wealth, and their internal resources, such as physical or intellectual capacities; it depends on the set of options from which they choose; and it depends on the degree of legal and non-legal constraints. The capacity to realize a given choice without constraint constitutes one opportunity.

A resourcist metric proposes to measure freedom by using proxies, that is, all-purpose and workable social primary goods. Rawls defines social primary goods as follows:

These are various social conditions and all-purpose means that are generally necessary to enable citizens adequately to develop and fully exercise their two moral powers, and to pursue their determinate conception of the good. (Rawls 2001, 57)

A useful proxy of freedom is income, for example, because, in contemporary market economies, it is an all-purpose good allowing one to buy a large variety of external goods, their price being fixed by their competitive value on the market. As Thomas Pogge notes, these goods should be all-purpose (primary goods) like income to avoid controversial perfectionist judgments that would arbitrarily restrict individuals' set of options and increase the likelihood of disagreements (Rawls 1971, 90–95; Pogge 2002, 208–212). As Samuel Arnold notes, these goods should also be as workable as possible (social goods) to avoid epistemic and implementation problems: “Primary goods, Rawls holds, must respect the limits of the practical: they must be amenable to objective measurement through a process that’s simple enough to be ‘workable’ for everyday politics” (Arnold 2012, 97; see also Pogge 2002, 217). This is a reason to focus on external goods like income which, contrary to internal goods, are more

---

5 On formulating the Rawlsian metric of justice in terms of freedom, see Van Parijs (1991, 225) and Sen (1992, 26–30).
workable for policymaking because they are easier to measure objectively. For instance, Rawls insists that “it is not self-respect as an attitude toward oneself but the social bases of self-respect that count as a primary good” and illustrates such social bases with “the institutional fact that citizens have equal basic rights” (Rawls 2001, 60).

Using all-purpose and workable goods as proxies for freedom solves two important problems in the measurement of freedom and opportunity sets. The first problem is normative: we must accommodate reasonable pluralism when trying to reach an agreement on the value of external goods like cars and bicycles. We cannot give intrinsic value to some goods rather than others because people’s preferences for different goods affect the value of these goods for them. If a person hates driving, she may not care for the opportunities created by a car and may prefer the ones created by a bicycle. This is why Arneson proposes a subjective metric of “opportunities for welfare” (Arneson 1989, 77; see also Van Parijs 1995, 48–51), which considers individuals’ actual preferences for specific goods when measuring the value of opportunity sets. Yet, one issue with relying on the subjective value of each specific good an opportunity set contains (for example, cars, bicycles, and so on) is that this would lead to disagreements about the value of the bundle of the least well-off, which in turn would lead to disagreements about what fair institutional arrangements are. While welfarist metrics make sense on a personal level, when we personally question what value goods have for us, a political conception of justice requires us to agree on what value these goods have for other people and, in particular, for the least well-off. Behind the veil of ignorance, we do not know who will end up in the least well-off position so we cannot base the evaluation of their bundle of goods on their actual preferences. This is why Rawls argues that “citizens’ appropriate shares of primary goods are not regarded as approximating to their good as specified by any particular comprehensive religious, philosophical, or moral doctrine”, and the reason to reject subjective understandings of the value of social primary goods is “to hold open the possibility of finding a public basis of justification supported by an overlapping consensus” (Rawls 2001, 60). When evaluating the share of resources attached to the lowest social position, we must adopt the perspective of a representative individual in the least well-off position and ask which combination of social primary goods would be rational for the least well-off to prefer (Rawls 1971, 94). In other words, we should represent ourselves as doing so on behalf of the least well-off, without knowing who they will be and with no information about
their preferences (for example, for cars or bicycles). This is one reason why Rawls recommends that: “For questions of social justice we should try to find some objective grounds for [interpersonal] comparisons, ones that men can recognize and agree to” (Rawls 1971, 90–91).\(^7\) Distributing all-purpose goods like income, that allow people to buy cars or bicycles as they prefer, is Rawls’ compromise to have some objective ground on which to build agreements on the value of the bundle of the least well-off, without making perfectionist judgments that could restrict individual options arbitrarily and lead to disagreements.

Using all-purpose and workable goods as proxies for freedom solves a second problem that is practical in nature. It is impossible to count the ‘number’ of opportunities or to measure the ‘physical size’ of the opportunity set opened by a given good, such as a bicycle. This is because any good opens an infinite number of unexpected opportunities, such as using bicycles to create contemporary art. Therefore, in practice, our lack of information about all potential uses of a good leads to incomplete rankings of opportunity sets because predicting every single thing that contemporary artists could do with a bicycle is not possible. One solution would be to stipulate that two opportunity sets are unequal only if one is a subset of the other. But this solution is inadequate because it also provides very incomplete rankings: it prevents any comparison between sets that do not overlap perfectly. If cars do things bicycles cannot do and vice versa, we cannot compare the opportunity sets opened by cars and bicycles.\(^8\) A more workable solution consists in using all-purpose goods like income to measure opportunities because it allows individuals to buy both cars and bicycles, thus avoiding the difficult task of comparing the opportunities opened by each specific good.

Considering individual preferences when measuring the value of opportunity sets, as in Arneson’s subjective metric, adds further workability problems. To begin, welfarist metrics require an account of individuals’ authentic preferences which raises epistemic problems: preferences may be ill-informed; people tend to discount future outcomes and misevaluate risks; psychological processes may lead individuals to develop expensive tastes or instead to adapt their preferences to their modest circumstances; interpersonal comparison of preference intensity is difficult; and

---


\(^8\) Van Parijs (1995, 21–22, 48–51) summarizes these standard difficulties in the measurement of opportunities.
people sometimes change preferences. Moreover, some point out the difficulty of constructing \textit{publicly verifiable} measures of individual welfare.\footnote{Several authors underline these common problems with welfarist metrics, including Rawls (1993, 187–190), Elster (1982, 219), Clayton and Williams (1999, 448–451), and Voorhoeve (2006, 266).} Finally, considering individual preferences may require making intrusive judgments about which preferences should count or not.\footnote{Anderson (1999, 310) raises this argument against assessing personal responsibility for our preferences.}

To sum up, measuring freedom by using all-purpose and workable proxies, such as social primary goods, is Rawls’ compromise to respect reasonable pluralism and workability while avoiding the normative and practical problems of subjective, welfarist metrics appealing to individual preferences. This provides some justification for a resourcist metric that includes all-purpose and workable social primary goods such as income and wealth, and the powers and prerogatives of positions of authority.

\textbf{Feature 2. Distributing Goods to Social Positions, Not to People}
Second, Rawls’ conception of justice does not actually measure how much social primary goods people have but instead what \textit{expectations} or \textit{prospects} they have over their complete life, given the distribution of social primary goods between \textit{social positions}. Indeed, Rawls explicitly rejects what he calls conceptions of ‘allocative justice’ which require that public institutions allocate social primary goods directly to individuals whose particular needs, desires, or preferences are known to us. He adopts instead what he calls a conception of ‘distributive justice’ which requires that institutions of the basic structure create \textit{social positions} to which social primary goods are attached, so that, if everyone follows the rules of social cooperation and takes up a position in a just cooperative scheme, the resulting distribution of resources will be just (Rawls 2001, 50–54; 1993, 283). The role of the basic structure is to distribute social primary goods, not directly to people, but to social positions.

The result is a \textit{two-step} system. In the \textit{first step}, social institutions create a cooperative scheme with social positions (for example, janitor, artist, teacher, lawyer) which combine a share of the burdens of cooperation (for example, labor) with a share of its benefits (the social primary goods). The difference principle applies to these social positions: the share of benefits attached to the lowest social position should be as high as possible. In the \textit{second step}, individuals can choose a social position under conditions of fair equality of opportunity (this explains why the...
two parts of Rawls’ second principle of justice are indissociable). In such a system, the principles of justice do not apply to the share of social primary goods that individuals end up having, but to the share of goods that are attached to the social position they can access (Rawls 1993, 283; 2001, 50–63, 175; Freeman 2007, 106–109, 125–127). What matters for social justice is the long-term expectations or prospects of individuals living in a society so organized. By maximizing the bundle of goods attached to the lowest social position, we maximize the life-long prospects of the least well-off and what they can expect even if they remain in the lowest social positions. As Philippe Van Parijs notes:

Properly understood, the difference principle is an opportunity-egalitarian principle [...] phrased in terms of expectations associated with social positions, rather than directly in terms of primary goods. (Van Parijs 2003, 214)

There are two kinds of reasons to reject an allocative conception and prefer a distributive one. The first kind concerns measurement issues discussed earlier: (a) reasonable pluralism would prevent agreements about principles to allocate goods directly to individuals according to their personal attributes such as preferences, needs, or merit, and (b) epistemic and feasibility problems involved in measuring what people prefer, need, or deserve would make such allocations unworkable. The two-step system of distributive justice, by contrast, avoids these problems because, once goods are distributed to social positions, individuals can choose the social position with the share of goods that best fits their preferences, including their preferences for occupation, resources, and work/leisure balance.

The second kind of reasons to prefer a distributive conception concerns the need for institutions of the basic structure to create the conditions for fair social cooperation to achieve justice over time (Rawls 2001, 50–61). For instance, making our access to social primary goods conditional on taking up a share of the burdens of cooperation may be important to create the incentives necessary to produce social primary goods in the first place. This is not because people should work to deserve a share of the goods: Rawls explicitly says that a conception of moral desert cannot be incorporated into a political conception of justice given the fact of reasonable pluralism (Rawls 2001, 73). Instead, this is because incentivizing people to work is instrumentally useful to produce the resources necessary to maximize the bundle accessible to the least well-off.
To sum up, in Rawls' conception of *distributive justice* (as opposed to *allocative justice*), the issue consists in finding an agreement on the value of the bundle of social primary goods attached to the *lowest social position* within a particular cooperative scheme. Achieving this would allow partners in the original position to compare different schemes and choose the one maximizing the share of goods attached to the lowest social position.

**Feature 3. Setting Aside Differences Between Persons**

Third, Amartya Sen argues that resourcist metrics of freedom are problematic because they neglect two elements that make persons different: their preferences and needs (Sen 1992, 85, 1999; Pogge 2002, 176–178). I explained why a resourcist approach does not take into account *preferences*, but freedom also depends on individual differences in *needs*, which depend on people's differing *abilities* to convert resources into opportunities. Since Rawls' conception of justice is primarily concerned with distributing burdens and benefits between fully cooperating members of society having abilities in the “normal range” (Rawls 2001, 175; Freeman 2007, 106–109), one could object that it wrongfully abstracts from injustices arising from illness or disability. To defend the Rawlsian account, one could underline that it does not ignore these injustices but simply breaks down the complex task of realizing justice into two steps.

In the *first step*, again, public institutions should secure a fair distribution of social primary goods to social positions (the indexing problem that concerns us in this article appears at this step). Succeeding at this step would already solve some injustices resulting from individual differences in needs and abilities. To begin, people that are able to participate in social cooperation but with less marketable talents will end up in the lowest social positions, but the bundle of goods attached to these positions will be as high as possible (Rawls 2001, 170–176; Pogge 2002, 196–204; Van Parijs 2003, 215; Daniels 2008, 46–58, 60–62). Moreover, existing differences in people's natural goods and internal abilities—such as the ones resulting from illness or disability—are often the result of current injustices. As Pogge underlines, in a fairer society, some of the social determinants of illness and disability would disappear (Pogge 2002, 186–188, 196–204; see also Daniels 2008, 79–88).

In the *second step*, once institutions have achieved a fair distribution of goods to social positions, they must secure access to these positions under fair equality of opportunity. This includes implementing policies to mitigate people's differences in needs and abilities, such as access to
health care and social insurance schemes, to compensate for life events, illness, and disability, and restore their capacity to participate in social cooperation or to convert their resources into opportunities (Rawls 2001, 175; Daniels 2008, 57). These are crucial issues of justice, but they are not the concern of the first step of the inquiry. I focus on the indexing problem arising when distributing social primary goods to social positions.

**Summary. A Simplified Indexing Problem**

Measuring freedom using an index of social primary goods raises problems: “One problem clearly is the construction of the index itself. How are the different primary social goods to be weighed?” (Rawls 1971, 93–94). Rawls simplifies this problem in various ways. First, he sets aside basic liberties to focus on goods varying in distribution:

Assuming that the two principles of justice are serially ordered, this problem is greatly simplified. The fundamental liberties are always equal […]; one does not need to balance these liberties and rights against other values. The primary social goods that vary in their distribution are the powers and prerogatives of authority, and income and wealth. (Rawls 1971, 93)

Moreover, Rawls’ resourcist metric focuses on all-purpose and workable proxies for freedom to respect reasonable pluralism. But at the same time, by focusing on social primary goods that everyone would rationally want more of to advance whatever ends they may have, it aims at fostering agreements on the value of the share attached to the lowest positions:

Greater intelligence, wealth and opportunity, for example, allow a person to achieve ends [they] could not rationally contemplate otherwise. […] While the persons in the original position do not know their conception of the good, they do know, I assume, that they prefer more rather than less primary goods. (Rawls 1971, 93)

In addition, Rawls’ distributive account of justice focuses on how institutions distribute social primary goods to social positions and then lets individuals choose the position satisfying their preferences. This avoids the need to take into account individual preferences when measuring the value of social primary goods. Finally, this account postpones issues regarding differences between persons’ abilities to a later step of realizing fair equality of opportunity. As Rawls summarizes:
The only index problem that concerns us is that for the least advantaged group. [...] It is unnecessary to define weights for the more favored positions in any detail [...]. If we know how the distribution of goods to the more favored affects the expectations of the most disfavored, this is sufficient. The index problem largely reduces, then, to that of weighting primary goods for the least advantaged. (Rawls 1971, 93–94)

Thus, Rawls simplifies the indexing problem by only focusing on the value of the bundle of goods attached to the lowest social positions.

III. THE INDEXING PROBLEM AND SUBSTITUTIONS

Despite the simplification described in the previous section, any metric of freedom based on more than one social primary good still faces this problem: Which combination of social primary goods can maximize the freedom of the least well-off? I now introduce important definitions to understand the substitution of social primary goods. Usually, the literature in economics adopts a metric of utility, that is, preference satisfaction, and defines substitutes and complements according to how goods contribute to the purpose of preference satisfaction. In the framework I propose below, I apply the same concepts to a metric of freedom, and define substitutes and complements according to how goods contribute to the purpose of increasing freedom.

Definitions 1. Substitutes versus Complements

Suppose that a metric of freedom $M$ depends on how much of certain goods one holds. For example, suppose that there are two goods $a$ and $b$. Substitutability arises when the amount of $M$ provided by $a$ may be replaced by the amount provided by $b$, and vice versa.

**Substitutes:** Given a metric $M$ affected by $a$ and $b$, goods are substitutes if an increase in the amount of $a$ ($x_a$) can compensate for a decrease in the amount of $b$ ($x_b$) while preserving the total value of the bundle.

Suppose a metric of freedom with only positive values and affected by a list of social primary goods. ‘Income and wealth’ and ‘economic powers’ could be considered substitutes if both provide independently valuable opportunities, that is, a person can use one good or the other to achieve the same purpose of improving their freedom. Instead, if goods are not substitutable at all, they are complements.
Complements: Given a metric $M$ affected by $a$ and $b$, goods are complements if an increase in the amount of $a$ ($x_a$) cannot compensate for a decrease in the amount of $b$ ($x_b$).

For example, electrical equipment and electricity have complementary functions because getting opportunities from electrical equipment requires electricity, and having no electrical equipment or no electricity suffices to prevent any opportunity from either of them. In other words, having one is necessary to make proper use of the other.

Definitions 2. Substitution Rates
In what follows, I consider a metric $M$ affected by two goods $a$ and $b$. Substitution rates are understood in absolute terms: if the goal is to preserve the total amount of $M$, the substitution rate $n$ refers to the absolute number of units of $a$ required to replace units of $b$.\(^{11}\)

If goods are perfect complements, each good is necessary but not sufficient to provide any unit of $M$. The level of $M$ is constrained by the good in the smallest quantity in the bundle ($x_a, x_b$). One way to formalize it is that their substitution rate $n$ is either 0 or infinite: $n = 0$ or $n = \infty$ (depending on how substitution rates are calculated). In such cases, perfect complements can be represented by a function of the form:

$$M(x_a, x_b) = \min \{x_a, x_b\}, \text{ for example: } M(4,2) = M(5,2) = 2$$

But for many $M$ and bundles ($x_a, x_b$), increasing the amount of $a$ still increases $M$ even with a fixed amount of $b$: $M(x_a, x_b) < M(x_a + \Delta x_a, x_b)$. For instance, a person does get more freedom if she gets a higher income even if her economic bargaining power remains the same, and vice versa. If so, it becomes possible to substitute some $a$ with some $b$ while preserving the value of the bundle. When substitutions are possible, substitution rates become important. We have to ask ourselves: For a given change in good $a$ ($\Delta x_a$) how does good $b$ have to change ($\Delta x_b$) to preserve the value of the bundle so that $M(x_a, x_b) = M(x_a + \Delta x_a, x_b + \Delta x_b)$?

If goods are perfect substitutes, an increase in the amount $a$ can compensate for a decrease in the amount $b$ while preserving the value of the bundle and the substitution rate is constant. For example, assuming a constant substitution rate $n = 1$ and fixed amounts $x_a$ and $x_b$ providing

\(^{11}\) The formalisation is generally following standard microeconomics but may diverge on occasion to simplify the explanation (see: Varian 2014, 38–41, 61–62).
comparable amounts of $M$, the value of the bundle $(x_a, x_b)$ is always the sum of $x_a$ and $x_b$, regardless of the distribution in the bundle: 1 additional unit of $a$ can compensate for 1 lost unit of $b$. In this case, perfect substitutes can be represented by the following function:

$$M(x_a, x_b) = x_a + x_b, \text{ for example: } M(1,1) = M(2,0) = 2$$

But for many $M$ and bundles $(x_a, x_b)$, substitution rates might not be constant. If so, the goods involved are only imperfect substitutes. There may be a variety of underlying reasons explaining why some goods are only imperfect substitutes, as I discuss later. If two goods are imperfect substitutes, the substitution rate will vary with the amounts $a$ and $b$ in the bundle. For example, a case of imperfect substitutes would be when one needs to raise the amount of $a$ by an increasing rate $n$ to compensate for the loss of each additional unit of $b$. One way to represent some kinds of imperfect substitutes would be by a function of the form:

$$M(x_a, x_b) = x_a x_b, \text{ for example: } M(4,4) = M \left( 5 \frac{1}{2}, 3 \right) = M(8,2) = M(16,1) = 16$$

Importantly, if goods are imperfect substitutes, the substitution rate $n$ is not constant but varies in absolute terms depending on the amount of the respective goods a person has in her bundle. This does not depend on individual preferences, which we have no information about, but on assumptions about the marginal returns of the different social primary goods in a bundle. For example, under the assumption of diminishing marginal returns, which I elaborate on below.

**Definitions 3. Diminishing Marginal Returns**

The account I presented assumes that freedom can only have positive values and is affected by several social primary goods. The core idea in this article is that social primary goods all have diminishing marginal returns, that is, the first units of each good provide more freedom than the last ones. If so, given specific assumptions specified below, they also have diminishing marginal substitution rates: other things equal, the less of one good there is in a bundle, the higher its substitution rate gets. In other words, other things equal, substituting one unit of $a$ with more $b$ is more costly when someone does not have much of $a$ and becomes less costly when someone has more of $a$. For example, if we take a balanced bundle, $(4,4)$, with freedom $M(4,4) = 4 \times 4 = 16$ as a reference, and if we want to
preserve its total value (16) then, for each additional unit of \( a \) that we want to replace with more of \( b \), the substitution cost will increase.

Figure 1 and figure 2 illustrate this idea. In figure 1, \( M \) is affected by an all-purpose good \( a \), with diminishing marginal returns. We assume that \( \mu a \) is a given unit change. As we can see, when someone does not have much of \( a \), one unit \( \mu a \) provides a substantial amount of \( M \), as reported on the vertical axis, but as the person gets more of \( a \), one additional unit \( \mu a \) provides less \( M \). This means that, other things equal, if the goal is to preserve the total amount of \( M \) in a given bundle \((x_a, x_b)\), replacing one unit \( \mu a \) by some of \( b \) will be more costly when someone has a low amount of \( a \) and less costly when someone has more of \( a \). In other words, other things equal and under conditions stated below, if good \( a \) has diminishing marginal returns, it also has diminishing marginal substitution rates.

In figure 2, the indifference curve represents all combinations of two goods \( a \) and \( b \) providing an identical amount of \( M \). Whenever goods \( a \) and \( b \) have diminishing marginal substitution rates, the curve is convex to the origin: when you have a balanced amount of \( a \) and \( b \), the substitution rate \( n \) is close to 1, but the more you depart from a balanced bundle, the costlier further substitutions become. When substitution rates rise quickly, the indifference curve is very convex. Evaluating how high substitution rates can get is an empirical issue.

The specific form of the substitution function depends on the kind of goods involved. The claim that whenever \( a \) has diminishing marginal returns, \( a \) also has diminishing marginal substitution rates is only true when the goods involved are assumed to be imperfect substitutes (not perfect substitutes) and all-purpose. Goods that are perfect substitutes, such as Dollars and Pounds, may both have diminishing marginal returns but not diminishing marginal substitution rates since they open the same opportunities. Goods that are complements and thus not substitutable such as electrical equipment and electricity may both have diminishing marginal returns, but because their functions are interdependent, they do not have diminishing marginal substitution rates: the value of the bundle always depends on the good that is in the lowest quantity. Finally, some goods may be substitutable but have substitution functions with irregular variations and thresholds, but in the case of all-purpose goods like income, giving access to a large variety of other goods, it is common and
plausible to assume that they are continuously divisible and that the marginal substitution rates will be the same (in absolute terms) regardless of the direction of exchange, which explains smooth substitution functions.

The core idea is that for imperfect substitutes and all-purpose goods, whenever they have diminishing marginal returns, they also have diminishing marginal substitution rates: the more you depart from a balanced bundle, the higher the substitution rate \( n \) gets.

With this framework in mind, answering the indexing problem raised by a resourcist metric of justice, and finding which combination of goods the least well-off should get, requires answering two related problems. The next section deals with the first problem, which consists in evaluating, in theory, which substitutions are acceptable. In other words, we need to evaluate the substitution rates of various social primary goods.

### IV. SOCIAL PRIMARY GOODS HAVE DIMINISHING MARGINAL SUBSTITUTION RATES

I now apply the framework presented above to the case of social primary goods. For the sake of simplicity, I focus on the goods that Rawls identifies as income and wealth and the powers and prerogatives of positions of authority since Rawls notes that “the primary social goods that vary in their distribution are the powers and prerogatives of authority, and income and wealth” (Rawls 1971, 93). I assume that these two goods are commensurable because they both contribute to people’s opportunities, but they do so in different ways. I argue that these goods are imperfect substitutes and all-purpose, and I give reasons suggesting that they have diminishing marginal returns and, therefore, diminishing marginal

---

**Figure 1**: Diminishing marginal returns

**Figure 2**: Indifference curve
substitution rates. To be clear, I do not want to argue for any specific list of social primary goods. I use these two goods simply to illustrate the framework I propose to think about the indexing problem.

First, what Rawls refers to as income and wealth allow for the exchange of various external goods on the market. In contemporary market economies at least, they are the most all-purpose and workable proxies of purchasing power. Distributing income and wealth helps to allocate other external goods. Individuals can use their income to purchase a variety of goods at prices set by competitive markets: scarce goods having higher prices and abundant ones having lower prices.¹²

Second, social institutions also create positions of authority and responsibility. These positions are attached to powers and prerogatives (Rawls 1993, 308; 2001, 58). Power in general can contribute to our opportunities by giving us the ability to bring about desired states of the world, some control over decisions affecting our lives, and protection from domination by others (Lovett 2010, 64–84). But one can understand this more specifically by mobilizing republican intuitions: social positions are attached to power over other people which can lead to domination (Lovett 2010, 74–84, 120). Frank Lovett defines power over as follows:

One person or group has power over another if the former has the ability to change what the latter would otherwise prefer to do—i.e., change the strategy the latter would otherwise select from their opportunity set. (Lovett 2010, 75)

Lovett proposes to measure the power of an individual or a group over another by measuring the degree to which the former can change what the latter might otherwise prefer to do, that is, can make a difference in the opportunities or strategies that the latter opts for. Because power is positional, what matters is not only our absolute level of power but the imbalance of power. There is domination when someone can exercise arbitrary power over others, uncontrolled by public rules or procedures (Lovett 2010, 78–84, 111, 120). The ‘powers and prerogatives of positions of authority’ could be understood as an all-purpose good. They are the

¹² In this article, income and wealth are understood in real terms which means their value is in real purchasing power. Distributing income for people to purchase external goods avoids the problem of agreeing on a method to compare the subjective value of all external goods given reasonable pluralism and differences in preferences. See Rawls (2001, 58–59), Van Parijs (1995, 41–45, 48–54), and Sen (1992, 28–30, 102–116).
social conditions securing the balance of power necessary to protect people from domination, which helps protecting their opportunities.

Now, income and wealth can be useful to get political influence or bargaining power. One could question whether ‘income and wealth’ and ‘powers and prerogatives’ are perfect substitutes. However, the opportunity sets they create do not perfectly overlap and having one does not necessarily allow getting the other. Indeed, the extent to which income and wealth can be converted into power over others depends on existing social institutions, as these examples illustrate. To begin, while income and wealth can sometimes buy political influence, democratic institutions can be designed to isolate politics from the influence of money by regulating lobbying practices and designing participatory mechanisms to allow people with fewer means to exercise political influence. These arrangements would limit the capacity to convert income and wealth into power over others or create a balance of power. Moreover, while an unconditional basic income could provide workers with more bargaining power in the labor market (Van Parijs 1995, 20–29, 32–38), a higher but conditional income does not always lead to more bargaining power. Firm managers can even pay higher wages to workers precisely to increase their power over them. Indeed, by paying workers above their marginal productivity, managers increase the cost of being fired and can use the risk of unemployment as a discipline device to make workers obey and accept difficult working conditions without complaint (Shapiro and Stiglitz 1984, 433). By contrast, unions are an alternative way to build bargaining power that is independent of income. Thus, income and wealth are not necessarily a perfect substitute for political influence or bargaining power.

The powers and prerogatives attached to positions of authority and responsibility can be understood differently, as the self-governing capacities required to make autonomous and intelligent use of our resources. This is a different, but equally important, social primary good. Indeed, social positions are also learning opportunities allowing us to develop important capacities required to make intelligent use of our resources. Rawls says that positions of authority and responsibility enable the development of the various “self-governing and social capacities of the self” (Rawls 1993, 308; 2001, 57–59; Arnold 2012, 98). Thus, Samuel Arnold understands this good as the social bases of self-governing capacities:

13 Large income and wealth inequalities create power inequalities and domination (Rawls 2001, 137–138). By contrast, some argue that a universal basic income could give all individuals more bargaining power in their economic relationships in the labor market by reducing exit costs (Van Parijs 1995, 20–29, 32–38; Lovett 2010, 196–200).
Offices and positions involving complex work cultivate the internal resources of intelligence and virtuosity [...] By “intelligence” I mean the ability to reason, to plan, to solve problems, to think abstractly, to comprehend ideas, and to learn. By “virtuosity” I mean skillfulness or cultivated aptitude [...]. Other things equal, a sharp and skillful person enjoys greater prospects of success across a wider range of endeavors than someone who is dull and incompetent. (Arnold 2012, 101, 114)

Note that for Rawls, the most important social primary good is what he calls the social bases of self-respect, providing us with a sense of our own value and the motivation and confidence necessary to achieve our goals. For him, the social bases of self-respect are: “Those aspects of basic institutions normally essential if citizens are to have a lively sense of their worth as a person and to be able to advance their ends with self-confidence” (Rawls 2001, 59; see also Rawls 1971, 440–446, 2001, 59; Freeman 2007, 113). Without a minimum of self-respect, nothing seems to have any value and one would abandon their goals. Yet, I set this good aside because, contrary to other goods, one could question whether the social bases of self-respect constitute a separate good. Rawls himself illustrates what such social bases are by referring to other social primary goods such as “the institutional fact that citizens have equal basic rights and the public recognition of the fact that everyone endorses the difference principle” (Rawls 2001, 60), which might be sufficient to secure self-respect.

Social primary goods may be complements up to some minimum level. A minimum of each social primary good seems necessary if only to secure the fair value of political liberties required by the first principle. For instance, the formal freedom of speech has no value if someone has no income to feed themselves, no protection from domination, or no intellectual capacity to meaningfully exercise free speech. A minimum of each social primary good may also be required to enjoy real opportunities because their functions are complementary at least up to some level. Income and wealth are useful for purchasing power, but without any protection from domination or any self-governing capacities, we may not be able to use our money in any useful way. But they arguably become imperfect substitutes after that: more income increases your freedom even if your protection from domination remains minimal. If social primary goods are substitutes, we must evaluate their substitution rates.

---

14 Anderson (1999, 317–319) discusses capabilities necessary for functioning as an equal citizen, for example.
I discuss reasons suggesting that Rawls’ social primary goods have diminishing marginal returns below. Yet, this claim may only be true for the least well-off who have quantities of goods within a certain range (in figure 3, they find themselves between ‘i.’ and ‘ii.’). First, extremely small amounts (between ‘0’ and ‘i.’) may be too small to be useful and initial units of these goods may have a flat or increasing marginal return. But in any social regime pretending to be just, the least well-off would have a bundle of goods that is above some minimum threshold. Second, in real societies, resources are scarce and unequally distributed. If a person has a lot of resources (above ‘ii.’), chances are that others have much fewer resources and the person with more resources can take advantage of the poverty of others, tilt the political process to their advantage, and get large benefits. This suggests that, above some threshold, the value of additional units may be increasing and not diminishing anymore. But the indexing problem I focus on concerns the bundle of the least well-off. Whatever happens in these specific cases (below ‘i.’ or above ‘ii.’), I only claim that social primary goods have diminishing marginal returns for the least well-off (between ‘i.’ and ‘ii.’). In this context, there are reasons suggesting that Rawls’ social primary goods do have diminishing marginal returns and, therefore, diminishing marginal substitution rates. Thus, the amount of freedom ($M$) opened by one social primary good ($a$) like income, arguably evolves as shown in figure 3.

One reason why the first units of social primary goods, like income, have more value than the last ones and, therefore, why their substitution rates can become very high is the clustering effect. Jonathan Wolff and Avner de-Shalit point out that disadvantages tend to cluster because they are interconnected and often reinforce one another (Wolff and de-Shalit 2007, 119–155; Daniels 2008, 61, 79–88). Indeed, low income is associated with various interconnected disadvantages, such as analphabetism, mental and physical illness, stress, obesity, child mortality, and teenage pregnancies; poverty also often comes with living in neighborhoods in which there is more violence and homicides, more drugs, higher rates of incarceration, and less social mobility (Wilkinson and Pickett 2010, 18–24). Therefore, when poor people get out of poverty, they tend to gain all at once a lot of new opportunities. By contrast, the same units of income create far fewer opportunities for a richer person because her basic needs
are already covered and she has already access to most basic consumer goods. Additional income only gives access to more consumer goods.lovett (2010, 143) claims that power has diminishing marginal returns in terms of opportunities. The clustering effect may again explain why. Indeed, the first units of power providing basic protection from domination can protect from the most gruesome forms of exploitation and violence. By contrast, getting more power when you are already protected against most forms of domination opens fewer additional opportunities.

Arnold (2012, 113–114) finally underlines that improving self-governing capacities is more important for disadvantaged people. Take the case of education and training. Basic primary education can provide all-purpose skills such as calculus and writing that are very useful in life and facilitate many opportunities for further education and employment. Basic education also indirectly impacts hygiene, health and life expectancy, the risk to engage in criminal activity, and teenagers’ reproductive choices. By contrast, achieving more advanced higher education or qualification mainly creates a few additional employment opportunities and perhaps some opportunities for intellectual satisfaction. Thus, early education has arguably more impact on crucial aspects of life. Of course, further research would be required to fully demonstrate these claims.

---

15 Empirical research suggests the diminishing marginal return of income in metrics like happiness or well-being too (Diener et al. 1993, 204; Veenhoven 1989, 15–18; Frey and Stutzer 2002, 90; Inglehart 2000, 219).

The clustering effect provides a good reason to believe that social primary goods have diminishing marginal returns. Under the assumptions specified in the previous section, we can conclude that they also have diminishing marginal substitution rates: other things equal, the less of a good there is in the bundle, the higher its substitution rate gets. Moreover, given how important the first units of each good are, the clustering effect suggests that the substitution rates of these goods can become very high, that is, their indifference curves are very convex.

V. SCARCITY CONSTRAINTS AND THE FEASIBILITY OF SUBSTITUTIONS

When trying to maximize the freedom of the least well-off, the first problem consisted in evaluating the substitution rate of social primary goods. Until now, I have analyzed their substitutability as if all acceptable substitutions were possible. Yet, some substitutions could be acceptable in theory but not feasible in practice because the substitution cost would be too high. Thus, answering the indexing problem requires solving a second problem: evaluating which acceptable substitutions are feasible in practice. In other words, we need to find which combination of social primary goods can maximize the freedom of the least well-off in the real world. This depends on how abundant or scarce social primary goods are.

The abundance or scarcity of social goods like income (contrary to natural goods like diamonds) depends on the social regime in place. If we consider only the social regimes that can pretend to realize a just society and maximize the freedom of the least well-off, these regimes would arguably be able to provide them with some amount of each social primary good. However, different regimes will provide a different combination of goods to the least well-off. The simplest illustration of this is that the relative economic productivity of different social regimes might affect the income attached to the lowest social positions.

If a regime were able to produce one good, like income, in absolute abundance, public institutions would always be able to afford the cost of substituting all other goods even if they had high substitution rates. However, if all goods are moderately scarce, public institutions have to choose the combination of social primary goods that can maximize the freedom of the least well-off within a limited set of feasible combinations. Moreover, the higher the substitution rates get, the more difficult it becomes for them to afford substitution costs.

Figure 4 illustrates how scarcity constraints restrict the set of feasible substitutions, that is, the 'domain of possibilities'. This depends on the
relative scarcity of goods, which depends in turn on the capacity of different social regimes to produce these goods. The production possibility curve delineates the domain of possibilities. Any combination of goods $a$ and $b$ that is outside the domain of possibilities is simply impossible, and combinations within are possible but do not maximize the value of the bundle because more $a$ or $b$ could be produced (in figure 4, I represent the production function that delineates the domain of possibilities as concave, instead of flat or convex. I discuss alternatives in figures 6 and 7). Given the range of combinations that would be acceptable in theory (on an indifference curve) and feasible in practice (within the domain of possibilities), the combination of two goods $a$ and $b$ that is feasible and maximizes the freedom of the least well-off is represented by the point in the domain of possibilities that intersects with the highest possible indifference curve. From this point, providing more $a$ to compensate for less $b$, or vice versa, would be acceptable, in theory, if it remains on the indifference curve, but it would not be feasible in practice because it would fall outside the domain of possibilities.

The scarcity of goods partly depends on how independent or dependent they are in production. Two goods are independent in production if making one good more available does not affect the availability of the other. By contrast, two goods are dependent in production if making one good more available requires making another good less available.\textsuperscript{17} For

\textsuperscript{17} Tomasi (2012, 189) and Arnold (2013, 394) refer to goods dependent in production as “rival”, not in the usual economic sense—that is, a good for which its consumption by
instance, *income and wealth* and *powers and prerogatives* may be goods dependent in production because policies required to improve the income of the least well-off may require reducing social protections preventing domination and vice versa. Market deregulation could foster free exchange and thus improve economic productivity and the income of the least well-off, but it might also increase capital mobility and competition which would reduce the bargaining power of low-skilled workers. On the contrary, implementing a basic income or democratizing workplaces could increase workers’ bargaining power and protection from domination, but these policies could negatively affect incentives to work and to invest, potentially reducing the overall productivity of the economy and, perhaps, reducing the income of the least well-off. Thus, these two social primary goods may be dependent in production, at least in some contexts, and this partly explains why they are moderately scarce.

One source of disagreement between ‘left’ and ‘right’ Rawlsians is how optimistic they are regarding the respective capacity of different social regimes to produce each social primary good. Like other liberals, John Tomasi considers that freedom should be the metric of justice. But Tomasi suggests that the freedom of the least well-off can be secured simply by giving them *income*. He reduces distributive justice to the distribution of income and does not care about other goods:

Free market fairness sees increases in income as holding out the promise of increasing the worth of the freedoms enjoyed by all citizens. […] Free market fairness interprets the difference principle as requiring that we increase the income of the least well-off. (Tomasi 2012, 190–191)

One way to reach this conclusion is to assume that more income can always compensate for other goods and that a Free Market System (FMS) can produce enough income to afford the cost of substituting all other social primary goods.

By contrast, Samuel Freeman explains that in a regime such as a Property-Owning Democracy (POD)—in which *income and wealth* are more dispersed but in which workers also have more of what Freeman calls one prevents its consumption by others—but in the sense that producing more of one good requires producing less of the other. Here, I use ‘dependent in production’.


19 Tomasi (2012, 189–191) and Arnold (2013, 394) disagree about how scarce or how dependent in production /rival social primary goods are.
economic powers, that is, more powers and prerogatives of authority—the value of the bundle of the least well-off is higher than in a welfare state capitalist regime (and a fortiori in Tomasi’s FMS) which only distributes income. For Freeman, the reason is that, despite potentially having a lower income, workers in POD have a larger share of other social primary goods:

In a “property-owning democracy”, workers’ share of economic powers and the bases of self-respect are greater than they are in a capitalist welfare state, since they have partial control over their working conditions and the management of production. In this regard, the index of primary goods of the least advantaged can exceed that of the least advantaged in the capitalist welfare state, even though the latter have greater income and wealth. (Freeman 2007, 113; see also Arnold 2013, 394; O’Neill 2008, 51)

One way to justify Freeman’s claim (without giving a priority to economic powers as he does in his later work, such as Freeman 2013, 31–32) is to demonstrate that welfare state capitalism or FMS are not productive enough to produce the amount of income required to afford the substitution cost of all other social primary goods.

VI. DISCUSSION: TOWARDS A SOLUTION TO RAWLS’ INDEXING PROBLEM
I have proposed a general framework to think clearly about the indexing problem within a liberal egalitarian, resourcist metric of distributive justice. If our goal is to identify which combination of social primary goods can maximize the freedom of the least well-off in the real world, evaluating their substitution rates is not sufficient. We also need broad empirical knowledge about the respective capacity of different social regimes to produce social primary goods and to distribute them to the least well-off. This task is beyond the scope of this article but to open the discussion and illustrate a potential upshot of this framework, I now consider a path towards solving Rawls' indexing problem. This suggests that maximizing the freedom of the least well-off is likely to require giving them access to a social position with a balanced combination of all social primary goods.

Compare the two theoretical regimes discussed above and their respective capacity to distribute income \((a)\) and power \((b)\) to the least well-off. In a POD, we assume that the workers occupying lower social positions have a balanced bundle with a good income and a substantial share of economic powers, through more bargaining power in their workplace which protects them from domination (Rawls 2001, 135–140; Freeman
2007, 219–235; 2013). By contrast, in Tomasi’s FMS, we assume that workers have an unbalanced bundle, with more income because the economy is more productive, but less economic powers because they have very little bargaining power in their workplace (Tomasi 2012, 190–191).

Each regime gives the least well-off access to a bundle of social primary goods that falls at the same time within the domain of possibilities and on an indifference curve of the least well-off. Between these two regimes, the one that maximizes the freedom of the least well-off is the one whose bundle falls on the highest indifference curve. Under this framework, to demonstrate that a regime such as a POD is superior to FMS, one would have to demonstrate that the combination of goods under POD falls on a higher indifference curve than it does under FMS.

Figure 5 illustrates the kind of demonstration required to conclude that POD is superior to FMS. We compare the two regimes by tracing the indifference curve running through the combination of income \((a)\) and power \((b)\) that they procure to the least well-off. If it can be demonstrated that the balanced bundle of goods under POD falls on a higher indifference curve than the unbalanced one under FMS, then POD is the regime best able to maximize the freedom of the least well-off. In this case, moving from POD to FMS and getting more income in exchange for less power would be bad for the least well-off. But this illustration relies on several assumptions. First, regarding indifference curves, if social primary goods have diminishing marginal returns and if their substitution rates can become very high, then indifference curves are very convex to the origin. Second, regarding the relative capacity of available social regimes to provide income and power to the least well-off, we assume that while FMS is more productive than POD, only some of the productivity gains benefit the least well-off. We also assume that producing more of one good at the expense of the other has an increasing opportunity cost, explaining the concave shape of the production possibility curve.

If these assumptions are verified, a higher indifference curve is reached more easily by providing a balanced bundle of goods, like in POD, than an unbalanced bundle with more income but less power, like in FMS. Indeed, even if we assumed that a social regime like FMS were more effective at producing income, it may not be productive enough to afford the high cost of a substantial substitution of power—taking as a reference the balanced bundle of goods of the least well-off under a POD. To illustrate, in figure 5, if our starting point is the bundle provided by FMS, reaching
the next indifference curve (as depicted) requires a large amount of additional income but only a small amount of additional power.

Defenders of FMS could question these assumptions regarding the production possibility curve. They could argue that, while POD gives the least well-off access to a balanced bundle of income and power, FMS is so productive that it can improve workers’ income sufficiently to offset lower economic powers. This is represented in figures 6 and 7. For simplicity, suppose that the production possibility curves are flat with a constant opportunity cost. In figure 6, FMS is slightly more productive than POD and the least well-off are better under POD, while in figure 7, FMS is significantly more productive and can substantially increase the income of the least well-off, thus reaching a higher indifference curve.

Comparing social regimes requires testing these assumptions about the shape and slope of the production possibility curve, defining the domain of possibilities. To begin, we should not assume too quickly that POD is less economically productive than FMS. Moreover, what matters in a Rawlsian framework is not the total productivity allowed by each social regime, that is, the ‘size of the pie’. What matters is the amount of income that each regime can give to the least well-off. This means that even if POD is less productive overall, it could nonetheless outperform FMS at maximizing the income of the least well-off. For POD to provide less income to the least well-off, it would have to be seriously less productive than FMS.

Yet, for the sake of argument, suppose that FMS does provide more income to the least well-off, although at the cost of reducing their economic powers. As shown in figure 6, if FMS is not sufficiently productive
to provide enough additional income to compensate workers for the loss of economic powers, POD would again remain better at maximizing the overall value of the bundle. As shown in figure 7, FMS would have to be exceptionally more productive than POD to be able to increase the income of the least well-off sufficiently to provide them with a combination of goods on a higher indifference curve than POD.

Determining whether this is true is an empirical issue beyond the scope of this article. But some literature suggests that a system like POD—despite involving more interventionist policies aiming at a significant redistribution of income and wealth and the improvement of workers’ share of economic powers by promoting unionization or workplace democracy—could be a viable social regime, productive enough to maximize the value of the bundle of the least well-off (O’Neill 2008; 2012, 75–100; 2020; Arnold 2013, 389–398; Freeman 2007, 219–235; 2013). This argument points towards a potential solution to Rawls’ indexing problem: if these assumptions are correct, maximizing the freedom of the least well-off is likely to require giving them access to a social position with a balanced combination of all social primary good.

**VII. Conclusion**

The contribution of this article is first and foremost to propose a framework to think about the indexing problem within a Rawlsian, resourcist metric of distributive justice. After introducing the main features of such a metric, I noted that solving the indexing problem requires addressing two related problems. The first consists in evaluating, *in theory*, under which conditions it is acceptable to substitute social primary goods, that
is, in evaluating their substitution rates. I argued that social primary goods have diminishing marginal returns and, therefore, diminishing marginal substitution rates: other things equal, the less of a good there is in the bundle, the higher its substitution rate gets. The second problem consists in evaluating which acceptable substitutions are feasible in practice. I argued that, because of scarcity constraints, some substitutions may be acceptable in theory but not feasible in practice, and the higher substitution rates get, the more difficult it is to afford substitutions.

While the article mainly aims at presenting this framework, I concluded by discussing a path towards solving Rawls’ indexing problem, which avoids giving any priority to some goods over others. Further empirical exploration is needed for a full demonstration, but I argued that plausible assumptions about social regimes and the feasibility of substitutions suggest that maximizing the freedom of the least well-off is likely to require giving them access to a social position with a balanced combination of all social primary goods.

If verified, this could open interesting lines of argument. One of them is a distributive response against arguments proposing to reduce distributive justice to improving the income of the least well-off, while avoiding the need to give any priority to some goods over others. Tomasi justifies a kind of Free Market System partly on the capacity of this regime to improve the income of the least well-off, but other social regimes may give the least well-off access to a more balanced bundle of all social primary goods. Tomasi might assume that income can compensate for all other social primary goods and that a Free Market System can produce enough income to afford the cost of substituting all other goods. But our discussion suggests that substitution costs might quickly become unaffordable. As a result, a social regime in which lower social positions are attached to a balanced bundle of social primary goods would be more likely to maximize the lifelong prospects of the least well-off. This could provide a distributive reason to prefer a regime like Property-Owning Democracy. Indeed, even if this regime proved to be relatively less productive, which could negatively affect the income of the least well-off, the overall value of their bundle could still remain higher if they have access to more powers and prerogatives of authority or self-governing capacities.
REFERENCES


Thomas Ferretti is a Fellow in philosophy at the London School of Economics in the United Kingdom. His research is interdisciplinary and focuses on three main areas: theories of distributive justice, fairness within economic organizations, and the ethics of artificial intelligence.

Contact e-mail: <t.ferretti@lse.ac.uk>
The Case of Stated Preferences and Social Well-Being Indices

SHIRI COHEN KAMINITZ  
Hebrew University of Jerusalem

IDDAN SONSINO  
University of Toronto

Abstract: This paper provides a real-world test case for how to approach contemporary preference aggregation procedures. We examine the method of using stated preferences (SP) to structure social well-being indices. The method has seen increasing popularity and interest, both in economists’ laboratory research and by governments and international institutions. SP offers a sophisticated aggregation of peoples’ preferences regarding social well-being aspects and, in this way, provides elegant and non-paternalistic techniques for deciding how to weigh and prioritize various potential aspects of social well-being (health, happiness, economic growth, etc.). However, this method also poses difficulties and limitations from broader political and philosophical perspectives. This paper comprehensively charts these difficulties and suggests that SP methods should be complemented with appropriate deliberation procedures. The paper bridges the distinct perspectives of economists and political theorists in order to make SP an attractive instrument in determining policy.

Keywords: well-being index, stated preferences, aggregation, deliberation, deliberative democracy

JEL Classification: A12, B40, C43, D70, I31

I. INTRODUCTION

Should policy goals be directly determined by citizens? This paper reviews new innovative techniques developed by economists that compose measurements of social well-being based on citizens’ stated preferences (hereafter: SP). We suggest that these techniques can be useful for public policy because they provide a more direct understanding of what citizens

AUTHORS’ NOTE: We would like to thank Avner de-Shalit, Ori Heffetz, Jack Vromen, and the anonymous referees for constructive comments on previous drafts. The paper was presented at the PPE conference, New Orleans 2019.
care about, but that they also raise significant problems that we need to be aware of when employing them. In short, we import new economic techniques into political and philosophical contexts and examine them from a broader perspective.

The new techniques we discuss can be viewed as a novel implementation of instruments of direct democracy. We argue that, like many measures of direct democracy, they are prone to serious pitfalls—if not complemented cautiously by other measures, in particular, measures of public deliberation. We argue that attempts to implement the SP techniques should incorporate deliberative methods in multiple stages. This perspective is different from that usually embraced by economists who tend to regard this method as part of a ‘social choice’ tradition, one which is less focused on political-theoretical considerations.

Following the growing sense that GDP growth is no longer a sufficient indicator for social progress and well-being, recent years have witnessed the development and implementation of both local and international composite indices supposed to represent social well-being (Fleurbaey and Blanchet 2013; Boarini, Kolev, and McGregor 2014; Stiglitz, Sen, and Fitoussi 2009). Within the process of indices’ structuring, various aspects of social well-being are selected (for example, education, health, life satisfaction, etc.), represented by indicators, and then combined within an index. This measurement of social well-being can then be adopted as a means for directing public policy (Bache and Reardon 2016).

This indexing process raises many questions concerning both the selection and combination of the indicators. The latter, the combination problem, often referred to as the ‘index problem’, is a pressing one if we refuse to be satisfied with a dashboard solution that leaves the separate indicators as they are instead of integrating them.¹ The index problem in our context asks: What relative weight should be given to each component that we believe social well-being encompasses?² This is a complex question. How can we determine whether health is more important than subjective happiness, and by what degree? Ascribing different weights to the indicators results in different indices and is therefore a process with

1 Fleurbaey and Blanchet (2013, 33) argue that the dashboard solution is problematic because it delivers too much information to be an efficient communication tool.
2 The term originates with Rawls (1971, 93–94), who talks about the weight each “primary good” should get when we assess the overall “disadvantage” the least well-off social group faces. For a review of Rawls’ approach to the problem and the major responses to his suggestions see Fleurbaey (2007).
political implications and significance (Decancq and Neumann 2016; Bache and Scott 2018).

SP methods for solving index problems have recently enjoyed particular attention from economists who have implemented and improved them (Benjamin, Heffetz, et al. 2014; Benjamin, Cooper, et al. 2017; Adler, Dolan, and Kavetsos 2017; Decancq, Van Ootegem, and Verhofstadt 2013; Decancq and Watson 2015). With regard to the index problem, there are instances of governments and international institutions adopting related mechanisms of consulting people about their preferences (Balestra, Boarini, and Tosetto 2018; Barrington-Leigh and Escande 2018, 910). These initiatives utilize up-to-date technologies and scientific techniques to collect SP data relatively easily and elegantly through internet platforms and computer software.

We can illustrate economists’ SP exercises by reference to the well-established Human Development Index (HDI), which is composed of three indicators: GDP per capita, life expectancy, and schooling. The index problem focuses on the procedure through which the three indicators are combined. Fleurbaey and Blanchet (2013, 14–15) point to the arbitrariness of the existing methods of combination within this index. The SP approach, in this case, suggests going to the people, that is to say, to a representative group of people who are directly consulted, through a survey, about their individual preferences regarding the components. Under particular assumptions, a social scientist may aggregate the data and elicit weights for each of the three components on a social level, and in this way resolve the index problem (Decancq and Watson 2015).

From an economist’s point of view this method is appealing. In this paper, however, we discuss the issue from broader political and philosophical perspectives and point to basic limitations in the SP method. These limitations, it will be argued, direct us to complement the method with various deliberation processes. Highlighting these limitations serves to enable the designing of a better process of aggregation and deliberation regarding the index problem. The paper points to the exact parts of the process where the two procedures of aggregation and deliberation should complement each other.

This leads us to an additional contribution our paper seeks to make. The problem of the relationship between aggregating citizens’ preferences and using deliberation techniques is a persistent and well-known...
one in political theory, social choice theory, and public policy.\(^3\) Should
group decisions be determined by individual votes or by achieving some
sort of consensus through discussion? If we want to integrate the two,
what is the way in which we should do so? Usually, these questions are
discussed theoretically from a general point of view, without looking at
the specific content of the choice problems encountered by the social
groups. Our paper shows that when looked upon from a careful examina-
tion of a particular case, we discover new specific difficulties in mere ag-
gregation and new specific functions for deliberation procedures that a
general discussion could not have identified. Our discussion thus contrib-
utes to the theoretical literature regarding aggregation and deliberation
by demonstrating how it can benefit from taking into account particular
types of political decisions.

The paper is structured as follows. The second section presents the
SP approach in the context of the index problem, highlighting its ad-
vantages. The third discusses the prominent (unique and non-unique) lim-
itations of SP methods in the social well-being indexing context from po-
litical and philosophical points of view. The fourth section discusses the
potential role of deliberation in coping with these limitations and pro-
vides the basic trajectories that, in this context, allow for deliberation and
preferences aggregation to complement each other to ensure a more fruit-
ful direction of well-being guided public policy. The fifth section provides
concluding remarks.

II. THE APPEAL OF SP METHODS IN DETERMINING THE INDEX PROBLEM

Scholars distinguish three categories of methods for solving the social
well-being index problem: expert-based weights, data-based weights, and
preference-based weights (Watson et al. 2019; see also Decancq and Lugo
2013; Balestra, Boarini, and Tosetto 2018). The latter attempts to elicit
individuals’ preferences regarding different potential elements of well-
being and aggregate them, under some explicit presumptions, in order to
establish social indexing.

In general, these preference-elicitation approaches to the index prob-
lem are considered attractive because they provide us with a practical ag-
gregation/weighing apparatus. Moreover, they are usually perceived, jus-
tifiably or not, as non-paternalistic and respectful of people’s own judg-
ments (in these respects, they are often considered superior to expert-

\(^3\) See for example Aldred (2004), Dryzek and List (2003), Elster (1983), Goodin (2003),
Knight and Johnson (1994), and Sunstein (1991).
based/data-based methods). These features have been emphasized recently in the economic and philosophical literature as leading criteria for policy oriented indexing processes (Fleurbaey 2012; Hyabron and Tiberius 2015; Alkire 2016, 625-629).

There are a few possible methods for eliciting such preferences (see for example Watson et al. 2019). Most notably, one can distinguish observational data—‘revealed preference’ techniques that look at behaviour as evidence of preferences (Bargain et al. 2013; Decoster and Haan 2015)—or use an external proxy to elicit preferences (for example, using SWB metrics as in Clark and Oswald 2002; Barrington-Leigh and Escande 2018). Alternatively, there are more direct techniques that use surveys to elicit preferences. Amongst the later, we could discern for instance willingness-to-pay surveys and also what are called here ‘SP surveys’ in which respondents directly rank/weight the alternative components. Thus, SP is but one method among various preference eliciting techniques and among various survey-based techniques (it is referred to here in this narrow sense).

Other preference-based methods and techniques have substantial advantages. For example, willingness-to-pay surveys have distinct strengths, such as their provision of a clear yardstick for comparisons and a potential for increased stability and consistency in the identified preferences (Streimikiene et al. 2019). However, while these techniques—and others—are suitable, SP methods have a special appeal in the social well-being index case, or so it seems. As we shall see, this is because this particular index problem might be interpreted in two distinct ways: as eliciting welfare/well-being or as determining social priorities. With the second interpretation, as a part of a political-democratic procedure, it may seem less appropriate to use monetized or other external benchmarks. For these reasons, recent research has tried to use surveys in which citizens are asked to directly rank well-being components one against the other in order to state their preferences explicitly (Benjamin, Heffetz, et al. 2014; Decancq, Van Ootegem, and Verhofstadt 2013; Adler, Dolan, and Kavetsos 2017; Decancq and Watson 2015).

---

4 According to this approach the weights of indicators in the index of social well-being “is determined by their explanatory power in statistical analyses of life satisfaction. [...] When explaining variation in life satisfaction using standard multiple regression techniques, important factors that can also be calculated at the individual level and therefore incorporated into the analysis will 'rise to the top' with a strong coefficient, i.e. weight” (Barrington-Leigh and Escande 2018, 911).
Our focus of attention on these SP methods is thus not because they are ‘better’ in eliciting preferences, but because they seem the most straightforward in their emphasis on respecting peoples’ attitudes (both as well-being subjects and as citizens with opinions). This feature makes it allegedly more resistant to the bulk of our criticisms. In any case, a great deal of the limitations we highlight here regarding SP procedures are relevant to the other methods of preference elicitation.

When investigating the SP method and its rationale, the following example may be of aid. Benjamin and co-authors (Benjamin, Heffetz, et al. 2014; Benjamin, Cooper, et al. 2017) advance one of the most thorough and comprehensive SP research programs, and so will be used here as a leading example. In Benjamin, Heffetz, et al. (2014), the authors composed a comprehensive list of 136 aspects of well-being, drawing from psychologists, philosophers, and economists, then asked 4,600 US survey respondents directly about their preferences, making pairwise comparisons between marginal changes in aspect $\alpha$ and aspect $\beta$.

By using a basic analogy between a social well-being index and a standard consumption-based-index, where well-being aspects substitute for commodities and services, the researchers elicit ordinal well-being functions from the preference data (analogous to ordinal utility functions). These researchers are thus using well-developed economic aggregation techniques in approaching the social-well-being index problem (Benjamin, Heffetz, et al. 2014, 2703–2707). From the aggregated data they attempt to pool a combined well-being index that reflects the relative importance each well-being aspect receives from the relevant population. Thereby, Benjamin and co-authors are able to estimate the well-being index $\sum_{j=1}^{J} \frac{\partial u(w)}{\partial w_j} w_j$ of a typical individual, where $w = (w_1, ..., w_J)$ represents the quantities of $J$ fundamental well-being aspects.

According to the survey conducted in this research, for instance, American respondents see the happiness of their families (ranked 2nd) and their ‘health’ (ranked 3rd) as more important for their personal well-being.

---

5 A PWB (personal well-being) index can be constructed using each individual’s MRS (marginal rates of substitution) for the aspects as weights (Benjamin, Cooper, et al. 2017, 82).

6 The theory of Benjamin and co-authors extrapolates the well-being index of a single individual, but for practical reasons they estimate the index of a certain ‘type’ of individual, pooling data from a large group of respondents while assuming they share the same marginal utilities. They warn that governments should only aggregate respondents whom there is good reason to believe share such indifference surface slopes but show empirical results suggesting differences between demographic groups may not be significant (Benjamin, Heffetz, et al. 2014, 2731–2732).
being than their physical safety and security (ranked 21st) or their material standard of living (ranked 98th).

Significantly, the researchers add to the survey another sort of question. While the first concerns personal well-being, this second concerns ‘policy-vote scenarios’, within which “the opening clause ‘Imagine you are making a personal decision’ is replaced with the clause ‘Imagine that you and everyone else in your nation are voting on a national policy issue” (Benjamin, Heffetz, et al. 2014, 2710).

Note how the two kinds of investigation refer to different kinds of judgements and preferences: the first in letting people reflect upon their own well-being (‘personal preferences’); the other as a ‘direct-democratic’ poll regarding what is best for society (eliciting ‘citizens preferences’). Whereas the former investigation is in common with economists’ preference elicitation investigations, which focus solely on representing and aggregating personal well-being preferences (as in Decancq and Watson 2015), the latter is in common with real-world initiatives such as the OECD better life initiative, directed to aggregating social priorities.8

An advocate of the SP method may perceive this dual-goals investigation as an advantage of the method. Within our examination of the SP aggregation method and its limitations, however, it is necessary to attentively separate the two kinds of tasks (eliciting personal preferences and political priorities). In the next section, therefore, we will discuss the two SP exercises separately.

In any case, the research as a whole, with its two different directions, is explicitly directed to guiding politicians (Benjamin, Heffetz, et al. 2014, 2699). This is by way of providing an elegant preference aggregation/weighting mechanism and a solution to the index problem, when the aspects ranked higher should be granted more weight. SP's greatest contribution in this context, in the eyes of its advocates, is when competing with expert-based procedures:

7 In Decancq and Watson (2015) people are asked about their preferences regarding the three components of the Human Development Index (HDI)—originally—average social indicators of education, health, and income. However, they are not asked about their preferences regarding the assessment of social well-being or development, but about the importance of the three components (personal health, personal education, personal income) to their own well-being.

8 With this web-tool, website visitors are invited to adjust weights for various well-being aspects, allowing the data to be recorded and aggregated to suggest an accepted index for different nationalities. Likewise, in the UK, the office of national statistics program conducted an attempt to construct a national consensus definition of national well-being using large-scale engagement of citizens via organized sessions and online commenting (Balestra, Boarini, and Tosetto 2018; Barrington-Leigh and Escande 2018, 910).
We believe it is more attractive to rely on what people’s own stated preferences suggest about what they themselves care about than to paternalistically rely on the opinions and introspections of “experts” (such as researchers and policymakers) regarding which aspects to track and how to weight them. (Benjamin, Heffetz, et al. 2014, 2702)

In the next section we examine this assertion from political and philosophical points of view and point to the essential (unique and non-unique) limitations of SP methods in this context.

III. LIMITATIONS OF THE SP APPROACH IN DETERMINING SOCIAL WELL-BEING

III.I. Stated Preferences and Social Welfare

Referring to SP as the appropriate method for eliciting weights for index components conceals implicit assumptions about the very purpose of the index. These presumptions should be carefully considered when specifying the role of preference aggregation. It is important, therefore, to spell out the assumptions about the index itself and what exactly it is supposed to represent.

The two different sorts of exercises of SP that were noticed (reflecting upon personal well-being and expressing social/political priorities) are in accordance with two different presumed goals for the index: tracking welfare and determining normative social/political priorities. In this sub-section we focus on the first and turn to the other in the next sub-section.

In the first case, stated preferences are taken as the appropriate method to extract or even to ‘track’ welfare. Two basic assumptions about the index are actually hidden here. First, it is assumed that social well-being is equivalent to social welfare; in other words, welfarism is assumed—the view that what matters ultimately and exclusively is the aggregation (by a specific rule) of personal ‘welfares’ (Adler 2019, chapter 1). According to this presumption ‘social well-being’ (the concept the index is meant to represent) and ‘social welfare’ are equivalent. Second, it is assumed that ‘preferences’ are the right information in determining personal welfare. Both presumptions could be debated, and hence could be adopted or rejected by political institutions and the public.

Regarding the first assumption, it could be the case that the role ascribed to the social well-being index is not to track personal welfares alone, but to track also some other conception that goes beyond welfarism, such as ‘social cohesion’, ‘social quality’, and so forth (although this is not common among economists, this conception of social well-being is acceptable sometimes by sociologists and philosophers; see Abbott and
Wallace 2012; Berger-Schmitt 2000; Wiland 2022). In these cases, it is not obvious that SP is the most relevant method.

Suppose that we do assume that the index should reflect aggregate personal ‘welfares’; still, as with other cases of social welfare functions we need to reflect upon the (personal) welfare/well-being conception and measure (Adler 2019). Choosing the measure depends on our conception of personal welfare. As a standard in the literature, we have three theories of personal well-being, or of how well an individual’s life is going (Adler 2019; Parfit 1986; Hausman 2011; Crisp 2017). Preference satisfaction is the view that personal well-being is a matter of preferences-realization; alternatively, the experientialist account is the view that personal well-being is a matter of mental experience, such as feelings, pleasures, satisfaction and so on; the third alternative, the objective account, posits particular objective qualities (such as accomplishments and friendships) as constituents of personal well-being, regardless of a person’s attitudes.

When we ascribe a welfarist role to the index and take personal well-being to be conceptually linked to preferences realization we have good reasons to favour the preference elicitation method in order to inform the social well-being index. Note how for economists the two presumptions about the index—welfarism and preference satisfaction as the relevant standard—come as natural. The two presumptions are a part of microeconomics’ raison d’être, and well-grounded in the economic paradigm (Hausman and McPherson 2006, chapter 8). Neither the public nor politicians, however, are obligated to take these presumptions as given. To use Manin’s terminology, these presumptions are not “evident, simple, and luminous”, and therefore, need to be deliberated in the strong sense of the term (Manin 1987, 347).

In case we deliberately favour the above conception of social well-being, it is still not obvious that we should prefer SP over other preference elicitation techniques. Thus, economists traditionally highlight difficulties concerning measurement and aggregation. A non-exhaustive list of problems would include interpersonal and intrapersonal comparisons (tackling preference heterogeneity—the possibility that different people have different preferences, a difference we would want to somehow reflect), stated preferences stability, and stated preference completeness. How far does the SP method allow these challenges to be overcome better than other techniques (Decancq and Nys 2021)? For example, willingness to pay surveys (alternative type of survey technique) and ‘equivalent income’ (a type of utility metrics), are regarded as capable of coping with
the above problems by using a monetized *numeraire* (Fleurbaey and Schokkaert 2013; Fleurbaey 2016).9

Benjamin, Heffetz, et al. (2014) acknowledge such difficulties and advise, therefore, utilizing other procedures such as SWB and money-metric approaches in particular, as complementary to their own. These methods complement the SP method by enabling the aggregation mechanism to be more precise (Benjamin, Heffetz, et al. 2014, 2732; Benjamin, Cooper, et al. 2017, 81).

This paper, however, highlights the possibility that part of the problems (among them completeness and stability) could be coped with through deliberation procedures.

III.II. Stated Preferences and Politics

We now turn to the other potential use of the method: employing SP for revealing policy priority preferences. This is an investigation in which the task is not to track personal welfare, but to represent the normative views of individuals regarding what is good for the society. Note how welfarism might constitute such a normative consideration in itself (i.e., one may presume that what is good for society *is* to promote aggregate welfare); however, this is not necessarily the case and there might be other goals to promote. For example, one could take sustainability as a granted weight in a social well-being index, even in cases in which it comes at the expense of current welfare (or at the expense of aggregated social welfare).

Focusing on SP for this political purpose, we should notice the alternative different scales of the indices: local and universal (McGregor 2018). An index may allow, on the one hand, the specification of the particular manner by which different societies address social well-being. With this local index problem, every society acquires *its own* solution to the index problem. Alternatively, it may (attempt to) provide a universal yardstick by which we can *compare* the well-being of different societies and nations, or the same society or nation across different times.11

---

9 Moreover, on a standard view, measuring well-being requires the construction of a von Neumann-Morgenstern utility function, which would require detecting the participants’ risk preferences (see Adler 2019, chapter 2). It is questionable whether SP is sufficient to establish such measurements of risk. These might significantly complicate the procedure (or direct attention to other techniques of preference elicitation).

10 Theoretically and empirically, it is reasonable and not uncommon for individuals to think on behalf of the group-as-a-whole (Tuomela 2013).

11 The project of social well-being indices in this context is often regarded a part of the ‘beyond GDP’ wave (Bache and Reardon 2016; Fleurbaey and Blanchet 2013; Stiglitz, Sen,
In accordance with the first goal—specifying a local index—some SP projects refer to the use of the method as an efficient tool for a policy maker when puzzling over criteria for redistribution and when deciding who is the worst-off within a society (Decancq, Fleurbaey, and Schokkaert 2015; Decancq and Schokkaert 2016). The second—universal—goal presents particular challenges for SP methods (for such an attempt, see Balestra, Boarini, and Tosetto 2018). In case we refer to the community of the whole world as one society, using SP in order to elicit the preferences of the world-wide population, the different (more controversial) preferences of the different sub-societies might end up marginalizing each other, marginalizing especially minorities. In any case, if we use SP to determine political priorities special attention should be given to the scope of the society in question.

Turning to another set of difficulties, the assumption that stated preferences is the right method to establish normative-political priorities is heavily debated in political theory. When pointing to the weaknesses of relying on preferences in political contexts in general, we can build on a prominent political-theoretical literature that explores the general pitfalls and shortcomings of resolving political issues by aggregating individuals’ preferences (see for example Elster 1983; Goodin 2003; Sunstein 1991). We shall briefly review five arguments raised in this literature that are relevant to our context and apply them to the SP method.

A significant first problem has to do with adaptive preferences (sometimes referred to as endogenous preferences). Preferences are adaptive when shaped by external circumstances with no intentional deliberation that may lead to their alteration (Bovens 1992; Elster 1983; Nussbaum 1999). We want to respect preferences because they generally reflect people’s autonomy, but adaptive preferences are not like that—they are not autonomous. Preferences about well-being aspects, both personal and social, may be surprisingly adaptive. One mechanism behind this is that

and Fitoussi 2009). It is also part of the broader and increasing phenomena of using indices and rankings in the domains of international relations and public policy (Cooley and Snyder 2015; Davis et al. 2012). In this context, an index is supposed to be capable of “being used to compare particular units of analysis, synchronically or over time” (Davis et al. 2012, 6).

12 This political goal resembles the motivation behind the literature concerning the measurement of poverty and disadvantage (Alkire 2016; Wolff and De-Shalit 2007).

13 It can be argued that the problem of marginalizing minorities is characteristic of the SP method in every resolution (it is also true in the local cases). Deliberation processes could mitigate this outcome to some extent.
many times people do not see the value of things until they get to engage with them to a certain degree. Some well-being surveys ask about the importance of arts and culture, but while many people who were exposed to complex forms of artistic endeavour (like visual arts or literature) find them an indispensable part of the good life, people in environments with less exposure to those elements will not in general claim to value them highly. Preferences regarding political rights may be adaptive as well. It is possible that people brought up in democracies find the participatory rights they enjoy a crucial part of social well-being while people who do not have them to begin with may find their contribution negligible, or even negative (see for example the recent case of Bhutan in Slater 2018).

A second problem is that stated preferences can avoid reflecting judgements about the common good and judgements about values. This can create a problem parallel in its structure to a collective action problem (Sunstein 1991). This issue can reiterate in the context of well-being surveys. This may happen, for example, for the following aspect in Benjamin, Heffetz, et al. (2014, 2715), which asks participants about: “The sense that you are making a difference, actively contributing to the well-being of other people and making the world a better place”. It may be that people will not define engaging in altruistic behaviour as important for social well-being when they think that people will generally not reciprocate the same sentiment.

Relatedly, it should not be taken as obvious that preferences successfully reflect values (Haybron and Tiberius 2015; Christiano 2008). Hence, some scholars distinguish A’s preferences—which may be whims that A finds no good reason to satisfy—from her values—a stronger and more distinct set of attitudes—in their attempt to define what A deliberately sees as good. One might answer that preferences could indeed reflect values, but we should consider, at the very least, complementing SP methods to bring preferences closer to reflecting people’s values. Social-deliberation methods could be of aid in this endeavour.

A third problem is that of illegitimate preferences: the possibility that some preferences that people will claim to have regarding well-being aspects are immoral. For example, it is possible that some individuals will

14 See for example Nagel et al. (2003) for some scientific evidence.
15 A recent paper demonstrated that while most Saudi men will not support the participation of their wife in the labour force, this preference is dependent upon a belief that most other Saudi men view female labour participation negatively. If they are told that this is false, Saudi men will shift their preferences and support their wife’s participation (Bursztyn, González, and Yanagizawa-Drott 2020).
hold a preference for an ethnically homogeneous society. It seems plausible to argue that higher-order moral considerations render such preferences problematic (Kymlicka 2002, 26–32). Benjamin, Heffetz, et al. (2014, 2703) address this worry and suggest that the hypothetical choices in abstract scenarios may elicit more deliberated and thoughtful states of mind (more, that is, than in ‘real life’ scenarios) that will wash out ‘emotional’ preferences. Although this is possible, still, there is no guarantee that this tendency will take place.

The common solution to these types of problem in the general case is ‘preference laundering’—we should not defer to the individual’s actual preferences but to their ‘purified’ or ‘laundered’ preferences (Goodin 1986; Griffin 1986, chapter 9; Hausman 2011; see Infante, Lecouteux, and Sugden 2016 for a recent survey of theoretical models that make use of this notion). Our discussion of deliberation methods in the next section provides some practical tools to bring about such ‘preference laundering’.

A fourth problem is that individual’s preferences (both personal and political) can be extremely unstable, perhaps little more than some form of primitive impression (Schumpeter 2003, 256–264; see Freeder, Lenz, and Turney 2018 for some recent evidence in a political context). If preference $a$ may be changed to $b$ and back in a short period of time, it seems that the case for respecting preference $a$ is undermined to a great extent. Moreover, in case it is unstable it cannot underpin a stable index.

Finally, it may be also the case that relying on preferences concerning social well-being priorities is a process that demands information and theorizing not available to most people. When trying to understand, for example, the place of inflation in individuals’ well-being (an aspect in the list of Benjamin, Heffetz, et al. 2014), it is hard to see how individuals can evaluate in a meaningful way its impact on their lives or the lives of their fellow citizens, its trade-off with other well-being aspects, etc. Goodin (2003, 54–57), for instance, provides basic principles regarding the question of which kinds of preferences are suspected of being ‘permissibly overridden’ in a democratic framework. Among them, he points to ‘uninformed preferences’ of various kinds.

Another example for this last difficulty is with comparing between (or specifying the trade-offs between) objective social circumstances and subjective attitudes. What is the meaning of comparing the level of freedom of speech, or rate of unemployment to a subjective average of life satisfaction in a society (a policy-vote scenario in Benjamin, Heffetz, et al. 2014)? The task of attentively addressing the relation between the two is
a complex one (and demonstrates a difficulty of parsing between instrumental and intrinsic valuations of goods).\textsuperscript{16} This kind of task within a survey violates a basic suggestion in methodological guidelines: to avoid giving the respondent too difficult a task to perform (Belson 1982, 389). It should be noted that the fact that respondents do provide their preferences regarding every two aspects does not guarantee their awareness of the issues behind the decision. There is a possibility that respondents use particular heuristics in order to avoid the complex decision-making behind the comparison.

The above problems can be mitigated, at least to some extent, by certain processes of deliberation (Manin 1987; Cohen 1996, 348–350; Sunstein 1991), both by experts and by the survey participants.

IV. EMBEDDING STATED PREFERENCES AND DELIBERATION

Focusing on the method of stated preference aggregation in the context of social well-being indexing allows explication of the concrete sorts of problematics that this aggregation procedure—taken as is—conceals. In this way it paves the way for complementing this procedure by various deliberation processes that could mitigate the problematic aspects—deliberation understood as a process through which individuals and groups justify decisions through giving reasons and exchanging information (Manin 1987; Cohen 1996; Gutmann and Thompson 2004).

Deliberation could be useful, in our case, as a process conducted by experts and by the survey participants. Deliberation processes are expected to allow for the SP aggregation process to become more focused and more appropriate for implementation. Deliberation processes have the dual potential to improve the inputs of the aggregation and to better the use of the output.

Thus, with the pitfalls and limitations of the SP methods borne in mind, the complementing deliberation procedures, so we suggest, should come at two stages of the process.

First, extensive deliberation processes should come prior to the SP exercise. Concrete dilemmas should be resolved. These preliminary processes could gain both from public deliberation and deliberation by experts. Among the deliberated issues, as clarified in the previous sections,

\textsuperscript{16} A similar observation is pointed to by Adler, Dolan, and Kavetsos (2017, 62), albeit they focus on personal well-being and not on political preferences. They call this problem a ‘hedonic forecast’ that may make the respondent believe that a higher level of the objective feature will in any case increase her level of subjective well-being.
it is necessary to specify the goal for the index. What concept exactly is it supposed to represent? Welfarism or something beyond? What is the scale of investigation (local or universal)? What is the purpose of the index (comparing between societies or between individuals within societies)? It is through resolving these issues that the exact role of aggregation is specified.

Secondly, group deliberation techniques should be better integrated into the SP exercise itself. Deliberation is directed in this stage of the process to strengthen self-reflection through conversation, to make preferences better reasoned and better informed, and to provide the opportunity to affect the judgments of others (Landa and Meirowitz 2009, 427). We should recall that in the political context of the exercises (eliciting ‘citizen preferences’), the main theoretical consideration that motivated the development and implementation of the SP method was to respect individuals and avoid the imposition of external standards on them. The political philosophy literature, however, reminds us that deliberative methods implement these values in a way that aggregating votes (as SP methods do) cannot (Cohen 1998, 2009; Dryzek 2007). While a citizen who answered the SP survey can know that her voice is one of the many voices that affected the exercise, the exercise’s results are still an opaque index that does not put forward any substantive justifications for the many policy implications it entails. Using discussion and debate as part of the process means that the citizens are actually presented with reasons underlying the policy choices and have richer opportunities to impact and challenge the policy decision. This implements a much more intuitive and convincing understanding of respecting individuals.

Moreover, deliberations will help us resolve many of the concrete issues we have identified with relying on stated preferences (in section III.II). Adaptive preferences, self-regarding instead of public-regarding preferences and values, illegitimate preferences, unstable preferences, and ill-informed preferences—could all, at least to some extent, be mitigated by deliberation between the survey participants. The process can stabilize the preferences (for a recent review of the empirical evidence see Kuyper 2018, 4–5). Being able to ask questions and to be exposed to arguments of other citizens will help participants to know what their fellow-citizens’ preferences are, and so avoid collective action problems and the neglect of values they would actually accept and acknowledge if only they knew more about them (like the value some people would find in publicly funded art or the public representation of marginalized minorities).
Political choices about well-being priorities are not individual independent preferences, they make assumptions about others, and not giving participants the opportunity to explore these assumptions generates distortions.

Furthermore, we believe that deliberation processes might be helpful even in the context of pure well-being tracking (the personal context). In the deliberation process, participants can be presented with ideas that they could have not taken into account when answering the survey in isolation. They can also make their preferences more robust and reflective through discussion and be encouraged to also deliberate intrapersonally—the more formal events in which interpersonal deliberation is conducted can in themselves encourage participants to think more seriously about the questions they are presented with, and so provide more hard-headed answers to them. These more thorough and comprehensive answers will provide a better, more consistent understanding of what actually underlies the participants’ well-being.

Think, for example, about someone who on a quick survey does not realize the importance certain economic opportunities or conditions provide to her, but after hearing more about their meaning, understands that they are a part of her own well-being. Deliberation, even in the tracking well-being case, may lead to ‘corrections’ in our index, and maybe also to ‘laundering’ preferences.

How should these participants’ deliberations take place in practice? Mini-publics, in which the chosen participants are convened to discuss the posed problem, are the paradigmatic institution that the literature on deliberative democracy directs us to (Ryan and Smith 2014). However, if we want a very comprehensive picture of how well-being is conceived among a large group of citizens, these might be challenging to arrange in practice. We can suggest two alternatives that have been developed in the recent literature. First, we can use online deliberation. Well-tailored forums designed by research institutions and government agencies can provide a platform for high-quality long-term discussion that will allow a large group of citizens to construct meaningful images of what they think about when they think about a life that is going well (for a useful review of practical and theoretical considerations regarding such institutional online deliberation, see Strandberg and Grönlund 2018). Alternatively, or in addition, we can use deliberative procedures that inform us about what kind of differences in answers we can expect if deliberative elements were to be incorporated along the whole body of respondents, such as
representative deliberative polling (Fishkin 1995). In these deliberative processes, citizens would be able to suggest well-being aspects of their own, present their own reasons for prioritizing certain elements, enter into a long-term process allowing them to reconsider such reasons, and to change their choices about the policy priorities. As part of these deliberation efforts, more extensive discussions with experts can be conducted, allowing the weighing in of expert knowledge that will not be incorporated in the original SP procedure.

A common concern is that while deliberation is an attractive ideal, it does not work in practice—and that the deliberative procedure creates polarization in the real-world group (Sunstein 2002), conformity pressures that can distort some of the more eccentric voices, and, more broadly, a discussion that does not revolve around respectful reasons-exchange and mutual learning (this might seem like a particular problem with online deliberations). However, as some empirical results have demonstrated, such low-quality deliberation, whose fundamental problem is probably the lack of a disposition to engage with peers in a way that might lead to a change in one’s opinion (Brennan and Landemore 2021, 223), is likely to occur only under specific conditions and contexts. In particular, such phenomena are likely to obstruct deliberation when individuals already have strong predispositions about the topic of deliberation, that they would seek to reinforce—for example, on topics already strongly politicized (Mercier and Landemore 2012). As defining well-being does not resonate right-out-of-the-gate with any salient public discourse, such patterns can be predicted to be less likely to occur in our context. Facilitated or moderated discussion, as those we are talking about will be, is also less likely to engender reduced quality (see Strandberg and Grönlund 2018, 369). Actually, a staple of successful democratic deliberation, the Icelandic constitutional process of 2010–2013 (Landemore 2015), revolved around giving citizens similar tasks of putting forward values they care about and deliberating on their relative importance.

The use of multiple deliberative procedures fits a framework that asserts that the question of the nature of social well-being is too complex to allow the provision of reliable answers through immediate responses; the answers should be produced in a long-term, multi-layered process. These methods of preferences elicitation do their job if we have

---

17 For a demonstration of an ambitious and successful process of deliberative polling regarding political priorities see Fishkin, Luskin, and Siu (2014).
18 For an elaborate case study demonstrating the potential contribution of facilitated online deliberation about political priorities see Albrecht (2006).
meaningful, well-formed preferences. But this is not the case with political preferences which require the processes of exploring and forming them.

Scientific scholarship has shown that deliberation processes have brought about changes in opinions and views, at least to some extent and in certain contexts (Ulbert and Risse 2005; Sulkin and Simon 2001; Fishkin and Luskin 2005). We stress again that the claim that deliberation will mitigate the political-theoretical problems of relying on stated preferences in this context is partly an empirical hypothesis that merits testing.\(^\text{19}\) It is not solely empirical, however. A part of the importance of deliberation cannot be proved by observation. The very legitimacy of the inputs is (by definition) strengthened through this process, as pointed out by Manin (1987, 352): “It is the process by which everyone’s will is formed that confers its legitimacy on the outcome, rather than the sum of already formed wills”.\(^\text{19}\)

Finally, acknowledging the problematic aspects of the SP aggregation (even in the deliberative mitigated cases), we could ascribe to the output of this aggregation procedure different roles within the broader decision-making process. As suggested in other contexts (Goodin and Dryzek 2006), the output, i.e., the index weights, could be used in various ways and not only for determining weights. It could also, for instance, inform an external process of determining weights and contribute to legitimizing particular indices.

V. CONCLUDING REMARKS

Stated preference methods in the context of the social well-being index problem carry significant strengths. However, they also have their limitations, especially from political and philosophical points of view. While the literature developing these techniques and the experiments implementing them have expanded in an accelerating manner, these limitations have not hitherto been systematically analysed. Their thorough analysis, we suggest here, allows us to address them in order to take better advantage of the SP innovations in the political context.

The supplemental suggestion advanced is that acknowledging the various strengths and pitfalls of the SP methodology clears the way for embedding deliberation procedures into the preferences aggregation

\(^{19}\) Thompson (2008) and Ryfe (2005) refer to the empirical challenges posed to deliberative-democratic ideas and techniques in many other contexts.
processes with which it provides us. Within these mixed procedures, a significant role will be reserved for the SP, albeit a limited one.

From a theoretical point of view, this paper can serve also as a lesson for other cases of embedding aggregation and deliberation in practical realms. Our paper shows that the question of how to combine aggregation and deliberation needs to be considered in particular contexts, rather than solely through overarching theories. The detailed analysis of a particular case could point to the unique difficulties it raises, together with the more general ones identified by political scientists and philosophers. Understanding the particular aggregation problem aids in tailoring better deliberation procedures that can help to overcome the problems.

This process is especially significant in cases where scientists point to new sophisticated aggregation methods. In such cases, even when aggregation methods are compelling, their implementation should be modified by external context-dependent considerations.

REFERENCES


**Shiri Cohen Kaminitz** is an assistant professor (lecturer) at the Hebrew University of Jerusalem, Political Science Department and PPE program. Her research interests include Political Theory, History of Economic thought, Utilitarianism, and Well-Being as a political agenda. She has
published in the journals: *Utilitas, History of Political Economy, Journal of Happiness Studies, Social Indicators Research, CRISPP.*
Contact e-mail: <shiri.cohen2@mail.huji.ac.il>

**Iddan Sonsino** is a political theory doctoral student at the University of Toronto. His research interests focus broadly on the intersection of political philosophy and political economy.
Contact e-mail: <iddan.sonsino@mail.utoronto.ca>
“The Hardest of All the Problems”: Hochman, Rodgers, and Buchanan on Pareto Optimal Redistribution

DANIEL KUEHN

The Urban Institute and George Washington University

Abstract: This paper provides a history of the development of Harold Hochman and James Rodgers’ (1969) theory of Pareto optimal redistribution, which modeled income transfers as a public good. Pareto optimal redistribution provided an economic efficiency case for redistribution policy. After reviewing the emergence of Pareto optimal redistribution at the University of Virginia and its elaboration at the Urban Institute in the early 1970s, the paper describes James M. Buchanan’s efforts to grapple with his colleague’s ideas in the context of public choice theory. Initially, Buchanan provided an even more expansive argument for redistribution than Hochman and Rodgers (1969). By the mid-1970s, though, Buchanan largely rejected his earlier approach to redistribution and theorized new criteria for redistribution that were narrower in scope.

Keywords: James M. Buchanan, Harold Hochman, James Rodgers, Pareto optimal redistribution, demogrant

JEL Classification: B21, B31, D63, D64

I. INTRODUCTION

Public choice theorists generally, and James M. Buchanan in particular, set a high bar for state action. Pareto optimality and unanimity are Buchanan’s ideals, tempered in practice by constitutional accommodations for the costs of decision making. These Paretian and Wicksellian hurdles are difficult enough to surmount for normal policy questions, but they become especially tricky in the case of income redistribution and inequality. Even a novice benevolent planner armed with uniform, twice differentiable utility functions could spin a tale justifying social welfare-improving redistribution policy, but how can a pure transfer possibly be a Pareto
improvement? As Buchanan and Tullock ([1962] 1999, 195) put it, “Pareto criteria can be drawn in for ordinary collective action, but they are useless here”. Conversing with Richard Musgrave decades later, Buchanan conceded that distributional questions were “the hardest of all the problems” facing public choice theorists (Buchanan and Musgrave 1999, 85).

This paper describes how a theory of Pareto optimal redistribution emerged out of the research agendas of the University of Virginia (UVA) Economics Department in the 1960s and continued to grow at the newly founded Urban Institute in the 1970s. The paper then describes how Pareto optimal redistribution was elaborated on and challenged by James M. Buchanan. Pareto optimal redistribution theory was developed in 1967 by Harold Hochman, a public finance professor at UVA and his student James Rodgers. The core of the theory is straightforward enough: redistribution can meet Paretian standards if utility functions are interdependent, and redistribution is a type of public good. The idea of redistribution as a public good had been stated at least as far back as Milton Friedman (1962, 191) in *Capitalism and Freedom* when he wrote that “I am distressed by the sight of poverty; I am benefited by its alleviation; but I am benefited equally whether I or someone else pays for its alleviation”.¹ Hochman and Rodgers did not mention Friedman’s insight in their paper, but their formal model of redistribution as a preference and a public good followed the same logic. They published their work in the *American Economic Review* in 1969 in a paper that inspired robust debate in the early 1970s. Hochman continued to develop the theory and apply it to practical policy questions with his public finance team at the Urban Institute in the early 1970s.

After describing the emergence of Pareto optimal redistribution, this paper reviews Buchanan’s grappling with the theory that his student (Rodgers) and colleague (Hochman) developed. Buchanan’s (1974a) initial public choice theoretic extensions of Pareto optimal redistribution justified more expansive income redistribution than even Hochman and Rodgers anticipated. However, by the mid-1970s Buchanan’s work on the economics of anarchy led him down another analytical path that allowed for a much narrower scope for redistribution that privileged minimizing distributional conflict and rent-seeking. This paper describes the contributions of UVA economics and public choice theory to the history of

¹ Drakopoulos (2012, 2016) reviews an even deeper history of interdependent preferences in economics, although the public goods aspect of redistribution highlighted by Friedman (1962) is what is essential to the Hochman and Rodgers (1969) theory.
theoretical criteria for redistribution policy and shows that the appropriate applications of theoretical standards to redistribution can be more complex and contestable than they first appear.

II. HAROLD HOCHMAN, JAMES RODGERS, AND THE DEVELOPMENT OF PARETO OPTIMAL REDISTRIBUTION

Harold (‘Hal’) Hochman was hired by the UVA Economics Department in 1965 after completing his PhD at Yale. By all accounts the students and faculty at UVA thought incredibly well of him (Pressman and Rodgers 2001). Buchanan (2006, 37) described how Hochman’s arrival “added to the excitement of the Thomas Jefferson Center”. Richard Wagner, a student of both Hochman and Buchanan at UVA, fondly recalls Hochman as providing the most “detailed and helpful” comments of any member of his dissertation committee.² Hochman was tapped to teach macroeconomics and public finance at UVA. The latter assignment in particular was an emphatic vote of confidence from a faculty that already boasted considerable expertise in the field.

Hochman and Rodgers began collaborating on their Pareto optimal redistribution paper when Rodgers was a student in Hochman’s spring 1967 public finance course. Their paper developed out of a classroom discussion on the weaknesses of Richard Musgrave’s (1959) text, Theory of Public Finance. Although the public choice tradition is often portrayed as counterposed to Musgravian public finance, Hochman saw great value in Musgrave’s text and used it in his class. He was frustrated, though, with Musgrave’s failure to provide the same economic efficiency justifications for the redistributive function of the state (the ‘distribution branch’) that he did for the allocative and stabilization branches. Musgrave’s allocative branch was concerned with the provision and allocation of goods, frequently focused on public goods but also the efficient allocation of private goods in the case of externalities. The stabilization branch focused on macroeconomic stability. Both the allocation and stabilization branches had clear efficiency criteria associated with them. The distributive branch, in contrast, required a “political process of decision making” (Musgrave 1959, 19) because market-based efficiency standards were unavailable. Musgrave (1959, 89) characterized the allocation branch as “more amenable to economic analysis” than the distribution branch, and “therefore it is proper for the economist to concentrate on this aspect”

² Richard Wagner, interview, December 1, 2021.
rather than redistribution. Hochman posed the problem as a challenge to his students: How could redistribution be justified on an efficiency basis? Rodgers agreed to work on the problem with Hochman, a collaboration that would generate their paper on Pareto optimal redistribution that was published in the *American Economic Review* in 1969.

The UVA Economics Department actively encouraged “research apprenticeships” like the one that generated the Hochman and Rodgers paper (Levy and Peart 2020, 60). Joint projects with professors frequently became a UVA student’s first peer reviewed publication and were a cornerstone of what Steven Medema (2011) calls the ‘creative community’ of public choice scholars. In fact, James Rodgers had multiple collaborations to choose from in the spring of 1967. While he was working with Hochman, Rodgers was also taking a class on cost and choice with Buchanan and had an appointment with him to discuss the possibility of co-authoring a paper. Finding Buchanan “too intimidating”, Rodgers skipped the appointment and the two never wrote together. Rodgers found Hochman more personally compatible, “very approachable, and easy to work with”, and they launched into their joint work revising Musgrave’s approach to redistribution.

Hochman and Rodgers (1969, 556) characterized their analysis as reconciling redistribution with “consumer sovereignty and an individualistic interpretation of the fiscal and political processes”. If individual utilities were interdependent, then efficiency would require transfers determined jointly by ‘transfer elasticities’, and by the initial distribution of income. The transfer elasticity was defined as the donor’s demand for an increase in the recipient’s income with respect to the differences in their incomes. Donors in the model cared about the absolute gap between their own income and the income of other persons, not the level of other persons’ incomes or utilities. In that sense, donors can be thought of as directly demanding reductions in inequality rather than having altruistic preferences that only indirectly relate to equity. The donor’s budget constraint for transfers had a slope of negative one and spanned half the gap between the donor and the recipient’s original income levels. The reason for

---

3 See Levy and Peart (2020, 60) for details on a proposal for “research apprenticeships” at the Thomas Jefferson Center to the Ford Foundation in 1960.

4 The class was built around material and texts that Buchanan would ultimately organize into the book *Cost and Choice* (Buchanan, 1969).

5 James Rodgers, interview, June 25, 2021.

6 For a detailed review of the related question of altruistic preferences, see Fontaine (2007, 2012).
spanning only half was that after half of the gap in income is redistributed, the donor and the recipient experience full equality, such that there would be no more possibility for difference-reducing transfers. Anti-social interdependent preferences, where individuals prefer to widen income gaps out of envy or malice could also be included in this framework, although Hochman and Rodgers (1969) ignore this case and later writers such as Harsanyi (1977) explicitly exclude it. Crucially, since donors value reductions in the gap between individual incomes and not their own donation per se, redistributive transfers are a public good. The benefit of another person’s donation to the poor is non-rivalrous for individuals who value reductions in income gaps, because all such individuals can experience the reduction in inequality at the same time. It is also non-excludable because once an income gap is reduced, the donor cannot keep other persons from enjoying the reduced inequality. Redistribution is therefore a public good for individuals who value equality of incomes.

After laying out the basic theory, Hochman and Rodgers demonstrated that the optimal form of income transfers depended on the magnitude of the transfer elasticity. An elasticity of zero made fixed sum transfers from each donor to each recipient optimal, while an elasticity of one made proportional transfers optimal. They compared modeled transfers under both of those extreme elasticity values to actual data on fiscal residuals to understand if the United States conformed to their model and if so, what it might imply about empirical transfer elasticities. Estimates of the distributional incidence of federal, state, and local taxes were relatively common in the empirical public finance literature by the late 1960s, but distributional studies of public expenditures and therefore total fiscal residuals were not. Irwin Gillespie (1965), a Musgrave student, provided Hochman and Rodgers with the expenditure estimates they needed. They compared Gillespie’s empirical data to two modeled fiscal residuals representing the extreme assumptions of their model: one case where the transfer elasticity was equal to zero, and another case where the transfer elasticity was equal to one (figure 1).

The first discrepancy between empirical fiscal residuals and modeled fiscal residuals that Hochman and Rodgers (1969) highlighted was that actual net transfers were disproportionately concentrated on income brackets that were lower income but not the poorest families. They

---

hypothesized that this may be due to differences in the political salience of rural compared to urban poverty, the lack of political power among the very poor, or even that it was just a statistical artifact. Stigler (1970), Tullock (1971), and Buchanan (1974a) subsequently showed great interest in the phenomenon of redistribution that benefited the near-poor but not the poorest Americans, which Stigler dubbed ‘Director’s Law’ after Aaron Director. However, none of them cite Hochman and Rodgers’s (1969) or Gillespie’s (1965) earlier analyses of the empirical regularity.

Figure 1: Comparison of Gillespie’s (1965) fiscal residuals with Hochman and Rodgers’ (1969) modeled fiscal residuals.

Notes: Fiscal incidence from Hochman and Rodgers (1969, 55; Table 7). Original income bracket labels from Gillespie (1965, 162; Table 11).
utility functions. Becker was working on two papers involving interdependent preferences at the time, one of which (“A Theory of Social Interactions”) was published in 1974 in the *Journal of Political Economy*. Becker’s papers were more broadly concerned with economists’ neglect of interdependent preferences than with the specific application to redistribution, but he appreciated the gap that Hochman and Rodgers were filling in their own public finance literature (Medema, 2015; Fontaine, 2007).

Hochman and Rodgers (1969) attracted three comments in the December 1970 issue of the *American Economic Review*. The first comment from Paul Meyer and J.J. Shipley (1970) criticized the model’s restrictive assumptions. A second, by Robert Goldfarb (1970), offered a specification of the utility function in the *n*-person version of the model that guaranteed Pareto optimal redistribution could not be accomplished by charity alone, but could be achieved through taxation. Meyer and Shipley (1970) were attempting to point out the theory’s vulnerabilities, while Goldfarb (1970) attempted to strengthen the theory by making government indispensible to redistribution. In their reply, Hochman and Rodgers (1970) found both comments to be either incorrect or relatively inconsequential for their core argument. The third commenter in the December 1970 issue was Richard Musgrave, whose public finance text provided the initial inspiration for Hochman and Rodgers. Musgrave’s comment was less technical than the other two, but ultimately more important for framing the discussion of Pareto optimal redistribution over the next decade.

Musgrave (1970) focused his comment on how Pareto optimal redistribution fit into his own typology of public finance, famously divided between the allocative, stabilization, and distributional branches. He pointed out that Hochman and Rodgers’ theory of redistribution itself depended on an initial distribution of income, and in that sense pertained to a ‘secondary redistribution’ that depended on a prior ‘primary redistribution’. Musgrave (1970) associated this primary redistribution with rules about inheritance, factors’ earning capacity, and market structure: the stuff of his allocative branch. He further pointed out that primary and secondary redistribution were intertwined, interacting with each other and difficult to untangle. Moreover, both were determined by the distribution of power and voting rights. Musgrave’s generally positive and elaborative comment on the Pareto optimal redistribution paper is notable because the initial purpose of the paper was to address a gap in

---

8 James Rodgers, interview, June 25, 2021.
Musgrave’s public finance. Rather than resisting Hochman and Rodgers’s criticism of the gap, Musgrave found their exercise valuable. It was also an important comment because Musgrave anticipated Buchanan’s elaboration on Pareto optimal redistribution, which explored the constitutional grounding for redistribution.

Hochman and Rodgers received criticism on the seminar circuit as well as in the journals. Rodgers recalls that when they presented the paper at Stanford, Michael Boskin dismissed the analysis as “empirically a tiny piece of the whole puzzle”. This was a criticism they heard repeatedly, despite the fact that Hochman and Rodgers did not argue Pareto efficiency was a justification for all redistributive policy. Although their paper was developed and published at a time when policymakers increasingly turned to economic efficiency standards and away from egalitarian ideals (Berman 2022), Hochman and Rodgers were primarily interested in the theoretical project of making Musgrave’s public finance more internally consistent rather than in promoting Pareto efficiency as an encompassing requirement for policy. If some redistribution was justifiable by Paretian standards then all the better, but they did not insist that the welfare state could only be judged on those grounds. This was a point that many contemporary critics failed to understand. For example, in his reply to Hochman and Rodgers (1969), Mishan (1972, 971) observed that “apparently economists have never been quite happy simply to propose transfers of income to the poor as a commendable aim of policy”, and instead prefer it if “the policy of equalizing transfer payments could be made to flow from some essential part of economics”. But Mishan’s critique misses the point. Clearly voters consider redistribution to be a commendable aim of policy and this is why redistributive policy is passed. Hochman and Rodgers wanted to analyze those voter preferences with the tools of economics rather than provide permission for those preferences or even worse, to dictate those preferences.

Hochman maintained his interest in Pareto optimal redistribution when he moved from UVA to the Urban Institute in 1970, where he led the Institute’s new public finance program. The Urban Institute was founded in Washington D.C. in 1968 and was intended to function like

---

9 James Rodgers, interview, June 25, 2021.
10 As an anonymous reviewer points out, this orientation is consistent with the view that Buchanan advocated (and often adhered to) that economists should provide citizens with an understanding of opportunities for mutually beneficial exchange rather than insisting on a course of policy. See Boettke and Candela (2018) for a recent elaboration of this characterization of Buchanan’s position.
the RAND Corporation, but for urban policy research instead of military research. Following the RAND model, the Urban Institute was largely dependent on federal research contracts, but its researchers did their work independent of the federal agencies that funded them (Finney 1968). The lion’s share of the research focused on evaluations of federal policies and programs. But just as economists at RAND pushed economic theory forward even as they worked on practical problems for the military, Hochman made theoretical contributions to the economics of redistribution alongside his colleagues who were working on program evaluations. Hochman supported the new public finance program at the Urban Institute with funding from the National Science Foundation (NSF) and the Ford Foundation.

Research on the distribution of income was “one of the Institute’s largest and most important programs” (Urban Institute 1972, 541) in its early years and was divided into three interrelated efforts. The first two efforts were computer simulation models: a static model for evaluating the impact of income maintenance programs and a dynamic model that incorporated a broader set of policy feedback loops affecting income distribution. The third research effort focused on “criteria for distributional policy” (Urban Institute 1972, 545) and was headed by Hochman. Criteria for distributional policy did not exclusively refer to Paretian standards, but it was this third effort that advanced the research on Pareto optimal redistribution. Although it was primarily theoretical, Hochman’s work was of considerable interest to the Urban Institute microsimulation teams. Harold W. Guthrie, the project lead for the income maintenance microsimulation project in the early 1970s, described Pareto optimal redistribution as a “promising attempt” at providing a badly needed theoretical criterion to guide federal redistribution policy (Guthrie, 1972, 418).

Many of the Urban Institute researchers working on criteria for redistribution were not primarily concerned with reconciling Musgrave’s fiscal branches or resolving Paretian paradoxes. Their work was practical, using the framework of interdependent utility functions to understand how transfer programs ought to be structured. George Peterson (1973), a member of Hochman’s public finance team, considered alternative...

---

11 See Weintraub (2017) for a detailed review of recent scholarship on the development of game theory and theories of rational choice at RAND and elsewhere, and their relationship to the Cold War.

12 The modern versions of the Urban Institute’s static and dynamic models for studying income distribution are TRIM3 and DYNASIM, respectively. For more details on the history of microsimulation at the Urban Institute, see Urban Institute (n.d.).
interdependent utility function specifications where donors cared about transfer recipients’ incomes and their leisure. In that case, the rate of substitution between labor and leisure was an important parameter for designing redistribution policy and balancing cash transfers with wage subsidies. Mitchell Polinsky, a student and coauthor of Musgrave’s at Harvard, became interested in Pareto optimal redistribution when he worked at the Urban Institute as a research assistant before starting graduate school. Polinsky published several papers expanding on the theory that grew out of his work as a research assistant (Polinsky 1971, 1972).

Urban Institute funding to work on Pareto optimal redistribution was also distributed to researchers at other universities, including John F. Johnston’s (1975) work at the University of Delaware and of course James Rodgers, by then a professor at the University of Pennsylvania. In one of his papers, Rodgers (1973) used Pareto optimal redistribution theory to contribute to the growing literature on the design of transfers, arguing that the relative preferability of cash or in-kind transfers depended on whether voters’ redistributive preferences were defined over transfer recipients’ welfare, income, or consumption. Within only a few years, Pareto optimal redistribution was featured in dissertations on welfare and redistribution (see, for example, Barmack 1975).

Most of Hochman’s funding while at the Urban Institute came from NSF, which was interested in basic research, including in the social sciences. More significant funding for the Urban Institute as a whole came from the Department of Health, Education, and Welfare, the Department of Labor, and the recently created Department of Housing and Urban Development. It was not easy for Hochman to interest those departments in theoretical economics, which had less obvious immediate relevance to program evaluation. Hochman ultimately struggled to support both his Pareto optimal redistribution work and the broader Urban Institute public finance team with regular infusions of contract and grant funding, and the team disbanded in 1973.¹³

Hochman and Rodgers’s work influenced other economists outside the circle of researchers supported by Urban Institute funds. Due to their UVA connections, Hochman and Rodgers also published on redistribution with Gordon Tullock, who liked their Pareto optimal redistribution paper and enjoyed talking with both Hochman and Rodgers about the

---

economics of redistribution, a long-standing area of interest of his.\textsuperscript{14} During this period, NSF also provided grant funding to Kenneth Boulding to study the economics of charity, interdependent preferences, and redistribution (Fontaine 2007), and Hochman and Rodgers even published an Urban Institute working paper on Pareto optimal redistribution in one of Boulding’s volumes (Hochman and Rodgers 1973). Kapteyn and Van Herwaarden (1980) noted that despite the interest in Pareto optimal redistribution, no one had attempted to empirically measure interdependent preferences and demand for redistribution. Citing Hochman and Rodgers (1969) as inspiration, Kapteyn and Van Herwaarden pioneered new approaches to this empirical problem.

Although Hochman and Rodgers’ perspective on redistribution generated considerable excitement, they were hardly the only economists analyzing the problem in the 1960s and 1970s, and arguably not even the most prominent. In the first third of the twentieth century, economists’ analysis of redistribution was primarily organized around the insight that diminishing marginal utility implied progressive taxation because an extra dollar had greater value to a person with lower income than a person with higher income. This style of argument, associated closely with Edgeworth, fell out of fashion in the 1930s and 1940s particularly following Robbins’ (1932) broadside against the utilitarian foundations of welfare economics.\textsuperscript{15}

Sandmo (2015, 51) describes a “comeback” for utilitarian approaches to redistribution in the postwar period in the work of Vickrey (1945) and Harsanyi (1955), who independently anticipated Rawls’s veil of ignorance. Vickery (1945, 329) described the optimal distribution of income in terms of an individual’s choice “were he asked which of various variants of the economy he would like to become a member of, assuming that once he selects a given economy with a given distribution of income he has an equal chance of landing in the shoes of each member of it”.

Another strand of analysis developed at the same time as the work of Hochman and Rodgers was James Mirrlees’ work on optimal tax theory. Optimal tax theory recognized that non-distortionary lump sum taxation was politically infeasible in most cases and asked what non-linear income

\textsuperscript{14} James Rodgers, interview, June 25, 2021. Although not considered in detail here, Tullock continued to be interested in the economics of charity and redistribution and published extensively on the subject, even more broadly than the Pareto optimal redistribution model. See especially Tullock (1981, 1997).

\textsuperscript{15} As Backhouse (2009) points out, Robbins is more accurately read as criticizing the stronger arguments of Hobson and Hawtrey than as a criticism of Pigou.
tax schedule would be optimal in a world where lump sum taxes are not an option. Analysis of the income tax as a redistributive policy lever brought trade-offs between equity and productive efficiency to the foreground and, to Mirrlees’ own surprise, implied low or even zero top marginal tax rates (Mirrlees 1971).

In contrast with all of these literatures, Hochman and Rodgers (1969) can be read as providing the basis for a realistic public choice model of redistribution. Vickery (1945), Harsanyi (1955), and Rawls (1971) provide some intuition for how citizens might choose a redistributive policy grounded in principles of justice or a social optimum, but their respective versions of the veil of ignorance introduce a wide gulf between the world of the model and real political systems. Mirrlees (1971) starts from more realistic assumptions, such as the restriction on lump sum taxes, but his is a social planner’s optimum rather than one that would reasonably emerge in a democratic society. Each literature has its insights and strengths, but it is clear how Hochman and Rodgers (1969) alternative perspective would have emerged from the milieu of the University of Virginia and also why Buchanan would want to build on it.

Broader ideological parallels between Hochman and Rodgers (1969) and Rawls also merit further comment. Pareto optimal redistribution theory emerged at the same ideological inflection point separating welfare state liberalism from neoliberalism as Rawls’ (1971) *A Theory of Justice*. Hochman and Rodgers therefore could be read (as Rawls is often read) as a justification for welfare state liberalism dressed up in the garb of social contract theory. Pareto optimal redistribution has the same aura of being a “last gasp of a dying welfarist ideology”, to quote Katrina Forrester (2019, xi) on Rawls. But just as Forrester (2019, xi) rejects this interpretation of Rawls in favor of the alternative view that Rawls was sorting out a mid-century “liberalism [which] was full of contradictions”, Hochman and Rodgers were also more driven to resolve the contradictions of the post-war liberal public finance tradition of Musgrave than in defending the welfare state directly. James M. Buchanan, meanwhile, was interested in and worried about the rise of the modern welfare state in addition to his contributions to theory. Buchanan is not usually identified as an important contributor to the literature on Pareto optimal redistribution, but he did consider the ideas of his former colleague and student carefully in the early 1970s. Buchanan’s reaction to their work served as a backdrop to his approach to redistribution through the rest of his career.
III. JAMES M. BUCHANAN AND PARETO OPTIMAL REDISTRIBUTION

Early in his career, Buchanan shared the orthodox view that transfers could not generally be justified or analyzed using Paretian standards. With his coauthor Gordon Tullock, Buchanan dipped his toes in the waters of redistribution, democratic choice, and Paretian standards in chapter 13 of the *Calculus of Consent* (a chapter that Buchanan drafted). The chapter assumed that pure transfers at the post-constitutional stage could not be Pareto improvements and argued in contrast that redistribution naturally followed from non-unanimity rules because a majority coalition would have an incentive to vote for transfers financed by the balance of the electorate. The *Calculus of Consent* notes that redistribution as insurance against the risk of future hardship might be agreeable at the constitutional stage of decision making, behind a veil of ignorance (Buchanan and Tullock [1962] 1999, 197). However, they argued that any post-constitutional redistribution (that is, redistribution after setting the rules governing policymaking) under democracy would tend to exceed these narrower insurance justifications (195).

After its publication, the Hochman and Rodgers (1969) paper was recognized by the community of public choice scholars as a work in their tradition. Mark Pauly (1970), for example, included it in his bibliography of recent public choice economics. Buchanan was familiar with Hochman and Rodgers’ work on Pareto optimal redistribution before it was published, since the acknowledgements in the paper indicate he provided comments on drafts. But Buchanan’s first public comment on the work came in March 1972, at a conference on redistribution and public choice organized by Hochman and the Urban Institute (published as Buchanan 1974b). Buchanan’s paper for the conference analyzed the level of government in a federal system that should undertake redistribution policy. He argued that Tiebout pressures (competitive pressures exerted on local government due to citizens ‘voting with their feet’) dramatically narrowed the redistributive scope of local government action but also ensured that local redistribution would be relatively efficient. National redistribution policy was less efficient and provided more scope for transfers because

16 James Rodgers has no recollection of Buchanan providing comments, so these likely came through Hochman rather than Rodgers (James Rodgers, interview, June 25, 2021). 17 James Rodgers also presented a paper at the Urban Institute conference and argued that a self-interest model did not adequately explain observed redistribution; some utility interdependence was required (Rodgers, 1974). As Richard Wagner’s (1976, 110) review of the conference volume playfully noted, this was a conclusion that “any self-interest model of intellectual choice would predict Rodgers to reach”.

VOLUME 15, ISSUE 1, SUMMER 2022 68
net taxpayers would have fewer viable opportunities to discipline the national government by voting with their feet.

Halfway through his analysis of federalism and redistribution at the Urban Institute, Buchanan raised the theory of Pareto optimal redistribution as one possible justification for redistribution at all levels of government. He argued that if national redistribution policy, with its tendency to inefficiency, could be grounded in interdependent utility functions that ensured Pareto optimality, then perhaps apparently inefficient national redistribution could be Pareto efficient. Buchanan harbored doubts about that possibility and threw himself at the mercy of the empirics. “If a case for federal-government or national-government activity is to be based on the grounds of strict utility interdependence”, Buchanan wrote, “evidence should be available to indicate that the sociocultural environment is such that the effective limits are, indeed, those determined by national boundaries” (Buchanan 1974b, 35). In other words, Buchanan raised no objections to the analytics of Pareto optimal redistribution, but he did demand evidence that individuals’ utility functions really extended to the welfare of individuals across the country, as opposed to locally.

A conference on public choice and redistribution at the Urban Institute, which had emerged as a hub for Pareto optimal redistribution research, would have been an appropriate opportunity for Buchanan to wrestle with the theory, but unfortunately his thoughts were not fully formed and articulated until later in 1972. In October, Buchanan developed his thoughts on Hochman and Rodgers’s work in much greater detail in a paper for a UVA conference in honor of Milton Friedman titled “The Political Economy of the Franchise in the Welfare State” (1974a). Buchanan’s interest in voting rules and public choice pulled him in a different direction in his evaluation of Pareto optimal redistribution. He began his talk by working out the marginal conditions for determining an optimal level of redistribution when utility functions were interdependent. This was a step that was not taken in the Hochman and Rodgers (1969) paper, although it was a natural extension of their work. Since the utility benefits of redistribution to a net transfer recipient are enjoyed by all net taxpayers, Buchanan (1974a) treated redistribution as a pure public good and derived the standard Samuelsonian condition for efficient government provision. The utility of one of \( n \) individual \( i \)'s was defined as a function of the income of individual \( i \) and individual \( j \),

\[
U^i = U^i(Y^i, Y^j)
\]
The \( n \) \( i \)’s were understood to be net donors to \( j \), the net transfer recipient. The Samuelsonian condition for the public transfer good was therefore,

\[
\sum_{i=1}^{n} \frac{u_{yi}}{u_{yi}} = 1
\]  

(2)

Where \( u_{yi} \) was the marginal utility of the \( n \) identical individual \( i \)’s, with respect to the income of a single individual \( j \). Since \( j \)’s income enters the utility of all \( i \)’s, transfers to \( j \) behave like a public good so the sum of the marginal utilities from the transfer equal the marginal cost of the transfer. Buchanan (1974a) pointed out that the problem with marginal condition (2) was that it excluded net transfer recipients from the political calculus. The marginal benefit of a transferred dollar to a transfer recipient was 1, so trivially

\[
\frac{u_{yi}}{u_{yi}} = 1
\]  

(3)

If these transfer recipients were included in the marginal condition for efficient provision of redistribution, the new Samuelsonian condition became,

\[
\sum_{i,j=1}^{n+1} \frac{u_{ij}}{u_{yi}} = 1
\]  

(4)

Buchanan (1974a) notes marginal condition (4) is not satisfied at the same level of redistribution as marginal condition (2). When transfer recipients are included, any redistribution exceeding the level defined by (2) will be a Pareto improvement and will not be rejected by transfer recipients. Because of the inconsistent Samuelsonian conditions, trying to pin down the efficient level of redistribution with interdependent utility functions proved a frustration. “This apparent paradox arises from the particular nature of pure income transfers”, Buchanan concludes, “and this nonuniqueness in outcomes makes the application of the standard tools difficult” (Buchanan 1974a, 55). The only sensible resolution for
Buchanan was to deny the franchise (that is, the right to vote) to net transfer recipients, at least on this particular policy question.\footnote{Buchanan (1974a) assumes voters vote directly on transfer policy. He provided no discussion of how to translate the disenfranchisement principle to representative government, where voters did not vote directly on policy.}

Although work on Pareto optimal redistribution at the Urban Institute in the early 1970s did not focus on the constitutional and voting rules developed in his analysis of Hochman and Rodgers (1969), Urban Institute researchers were aware of and debated the theoretical relationship between redistribution and voting rules that Buchanan raised. Richard Wagner, a Senior Research Associate at the Urban Institute from 1972 to 1973 after studying under both Buchanan and Hochman at UVA, recalls that members of Hochman's team discussed the relevance of A.V. Dicey's (1905) \textit{Lectures on the Relation between Law and Public Opinion}, which decried the enfranchisement of recipients of poor relief in England (see especially Lecture VIII).\footnote{Richard Wagner, interview, December 1, 2021.}

Ultimately, Buchanan (1974a) rejected his own proposal for the restriction of the franchise in democracies where it had already been expanded, not because a restricted franchise was analytically unpersuasive to him, but because it was impractical and "a shift [away from universal suffrage] would harm some members of the community" and therefore be inconsistent with Paretian standards (Buchanan 1974a, 71). This argument against franchise restrictions is interesting in light of Levy and Peart’s (2020) work on Buchanan’s views on the economics of natural equals. Here Buchanan is \textit{not} opposing franchise restrictions because natural equals deserve an equal vote. In theory, he sees franchise restrictions as justified. Instead, he is opposing franchise restrictions because it would not be a Pareto improvement in the U.S. case where the franchise was already (essentially) universal. But this acceptance of universal suffrage combined with the paradox posed by the Samuelsonian conditions under universal suffrage painted the public choice analysis of Pareto optimal redistribution into a corner. To get out of that corner, Buchanan made a remarkable observation that is worth quoting at length:

> Because political transfers seem to violate the standard Pareto norms, we have tended to opt out of any discussion and say that nothing further can be constructed on the basis of individual evaluations. I now propose to modify this stance, and to do so in a substantial sense. Failure to carry out income transfer policy in a political setting such as that sketched out in the models above would represent a
fundamental shift in the real assignment of rights, a shift in favor of those with the nominally higher incomes. Once the franchise has been extended to all adults, and once the constitution has allowed income transfers to take place collectively, the formal act of transfer becomes fully predictable from positive economic analysis. The basic property right inheres in the voting franchise, and the economic value of this franchise reflects the measure of the contingent claim to the incomes and wealth nominally imputed to individuals in the whole community. (Buchanan 1974a, 70)

This was, in Buchanan's own words, an acknowledgement that the prior understanding of Pareto optimality and redistribution “has been seriously deficient” (Buchanan 1974a, 62) because it had not considered the broader political context for redistributive policy. The first claim of the quoted passage, that ‘we have tended to opt out of any discussion’ of Paretian transfers was clearly untrue of Hochman, Rodgers, and their collaborators and elaborators, but it was a fair description of most of the rest of the profession. Rather than proceed in older ways of thinking, Buchanan argued that the constitutional decision to expand suffrage implied a contingent claim to the whole political community’s income and wealth. Buchanan argued that if redistribution was on the table as a constitutionally admissible policy choice in a given constitutional order, then clearly the constitutional order recognized that voters had a ‘contingent claim’ to all wealth and income that could potentially be redistributed (even if actual policy did not redistribute all potentially redistributable wealth). Buchanan’s analysis essentially collapsed Musgrave’s (1970) distinction between primary and secondary redistribution, opening the door to a wider range of redistributive opportunities than even Hochman and Rodgers (1969) envisioned.

In the discussion that followed Buchanan’s paper at the Friedman conference, Bill Niskanen confessed that Buchanan’s proposition about claims to social wealth as the “economic value” of the franchise was “one that I am queasy about”. Responding to Niskanen, Buchanan conceded that “it does suggest a tremendous broadening of this whole concept of Pareto optimality”. Almost as if to reassure his queasy audience, Buchanan then admitted “I don’t claim to have worked out all these implications” (Buchanan 1974a, 85–86).

Buchanan’s paper and the transcript of the discussion that followed his presentation leaves the reader with a palpable sense of tension, incompleteness, and lack of direction. The conferees were intrigued by the analysis but uneasy, and Buchanan seemed to share their hesitance.
“The Political Economy of Franchise in the Welfare State” was Buchanan’s most thoroughly developed analysis of Pareto optimal redistribution, and it is also one of the few cases where he treated the theory as relevant to the constitutional stage of decision making. In most other analyses where he cites Hochman and Rodgers (1969) directly, he cites them only briefly as one consideration of voters in their post-constitutional decision making—one possible reason among many to vote for redistribution in the post-constitutional policymaking period (see, for example, Buchanan and Bush 1974; Brennan and Buchanan 1980).

Although Buchanan never analyzed the Hochman and Rodgers (1969) model in any detail after the two 1972 conferences, the queasy idea that voters had a ‘contingent claim’ to all the wealth of society was an idea that he did not shake off lightly or immediately. In May 1974, at a conference on markets and morals at the Seattle Research Center of the Battelle Memorial Institute, Buchanan asserted the point even more definitively that “the individual’s right to vote, held on an equal basis with other people, implies an ultimate equality in a political claim to all wealth in society” (Buchanan 1977, 70). This was an extraordinary claim, and one Buchanan took seriously.

It is perhaps unsurprising that Buchanan took Hochman and Rodgers’s interdependent utility functions so seriously in the first half of the 1970s, since this is also the period when he was engaging in a deeper reading of ‘the full Adam Smith’, including Smith’s views on sympathy in The Theory of Moral Sentiments. Smith’s analysis of sympathy anticipates both the versions of the impartial spectator concept developed further in Vickery (1945), Harsanyi (1955), and Rawls (1971) and the interdependent utility functions of Hochman and Rodgers (1969) and Becker (1974). Buchanan found the arguments of The Theory of Moral Sentiments appealing and cited them in his analysis of Rawls (Buchanan 1976b) and to counter arguments by Gordon Tullock.

---

20 Buchanan’s writings can be notoriously sparing in citations, and remarkably he does not even cite this point to his own paper, “The Political Economy of Franchise in the Welfare State”, although that is clearly the origin of the argument.

21 I owe this insight to and borrow the nice turn of phrase ‘the full Adam Smith’ from an anonymous reviewer.

22 See Levy and Peart (2020, 40) for a detailed discussion of Buchanan and The Theory of Moral Sentiments, including a transcription of Buchanan’s 1971 letter to Tullock, which cited the book in an argument against Tullock’s theory of revolution.
IV. Buchanan’s Demogrant and the Narrowing Aperture of Constitutional Redistribution

By 1972 Buchanan had followed the thread of Pareto optimal redistribution to the startling conclusion that universal suffrage implied a claim to the wealth of everyone living under the constitutional order. Actual patterns of redistribution required democratic consensus, but the franchise implied a collective right to redistribute. A nagging question for Buchanan was how a constitutional order that opened the door to such radical redistribution could ever be agreed to in the first place. To answer these types of questions, Buchanan drew on the emerging economics of anarchy research program at Virginia Tech.²³ The economics of anarchy was in many ways the economics of conflict, where theft and appropriation by physical force was always a threat.²⁴ The question was, how did a stable constitutional order emerge from that fundamental conflict? As Buchanan (1992, 101) later described, “the research problem seemed to be to explain and understand the emergence of order out of anarchy rather than to grasp the meaning of a stable order that already existed”. This reorientation proved well suited to avoiding some of the undesirable conclusions that Buchanan drew from the theory of Pareto optimal redistribution at the 1972 conferences discussed above.

Winston Bush, the star of the group at Virginia Tech working on anarchy, was thinking about redistribution, stable orders, and anarchy alongside Buchanan in the early 1970s. Bush developed a model of distribution with an anarchic starting point where property rights were mutually agreed upon to secure wealth against predation, but at the loss of some initial wealth that weaker parties could not defend in the absence of property rights. The model did not imply the sort of claims to social wealth derived by Buchanan (1974a), and Bush (1972, 15) dismissed interdependent preferences specifically as “questionable”. Instead, Bush (1972) envisioned a stable order of balanced conflict emerging out of anarchy. Geoffrey Brennan (1973) made a similar argument at approximately the same time, which he called the ‘non-altruistic dimension’ of Pareto optimal redistribution. Redistribution in this post-anarchic constitutional context of balanced conflict took Buchanan down a different analytic path in his own

²³ See Buchanan (1972) for a forthright statement about the limits of traditional social contract theory and the prospects of the economics of anarchy to extend constitutional economics.

²⁴ See Bush and Mayer (1974) for a classic Virginia Tech style model of the economics of anarchy, equilibrium conditions, and ‘orderly anarchy’.
Buchanan (1975a) argued that indulging demand for redistribution can maintain the constitutional peace and provide a better outcome for net donors than revolution and anarchy.  

Hobbesian anarchy became a principal lens through which Buchanan analyzed questions of inequality and redistribution. He scrutinized Rawls’ original position (Buchanan 1976a) and Nozick’s entitlement theory (Buchanan 1975c) using the Hobbesian model. Although Buchanan’s (1975b) Samaritan’s dilemma paper did not explicitly draw on the economics of anarchy, it was motivated by similar concerns about distributional conflict. In these papers, Buchanan criticized the new contractarians’ unrealistic conflict-free original state, but he did not explore what type of redistribution was implied by the economics of anarchy. However, in a recent paper, Lehto and Meadowcroft (2021) persuasively argue that Buchanan’s demogrant proposal represents his principal resolution to the prisoner’s dilemma posed by the economics of anarchy.

The demogrant, more commonly known as a guaranteed income, was developed by Buchanan (1997), Brennan and Buchanan ([1985] 2000), and Buchanan and Congleton ([1998] 2003) as a uniform per capita transfer payment made to every citizen, financed by a flat tax. Buchanan argued that because the demogrant and flat tax offered no scope for preferential treatment they could eliminate the distributional conflict implied by the economics of anarchy (and for that matter, majoritarianism). This justification for the demogrant was a major break with Buchanan’s earlier reactions to Pareto optimal redistribution theory. Brennan and Buchanan ([1985] 2000, 143) justified a flat tax not by arguing that it meets Paretian standards but because it allegedly constrains coalitions of high-income individuals from “organizing large scale transfers to themselves”. Buchanan and Congleton ([1998] 2003) called this a ‘political efficiency’ argument for a flat tax, in contrast with more traditional conceptualizations of efficiency invoked by Hochman and Rodgers.

Buchanan and his coauthors stress the importance of a demogrant that is built into the constitution, to guard against opportunistic or

25 See Petach (forthcoming) for a careful discussion of Buchanan’s views on these points.
26 The demogrant proposal was primarily developed in Buchanan (1997), Brennan and Buchanan ([1985] 2000), and Buchanan and Congleton ([1998] 2003).
27 My qualifier that it ‘allegedly’ constrains self-interested coalitions is deliberate. Brennan and Buchanan ([1985] 2000) provide no real argument for why the aspirations of political coalitions would be eliminated by lump sum payments and flat taxes. This is an underappreciated criticism that I think applies to most invocations of the generality norm by Buchanan, Hayek, and others. Generality may or may not be desirable, but it surely offers no guarantees against opportunism.
cycling adjustments to redistribution policy. Lehto and Meadowcroft (2021, 161) call this design decision “welfare without rent seeking”. Thomas Romer’s (1988) discussion of redistribution in his Journal of Economic Perspectives review of Buchanan’s contributions, following the 1986 Nobel Prize, is almost entirely concerned with this insistence on embedding redistributive decisions at the constitutional stage.

Buchanan had not always supported demogrants or other universal basic income schemes. In his 1968 paper, “What Kind of Redistribution Do We Want?”, he dismissed Milton Friedman’s negative income tax proposal as incompatible with voter preferences. Voters do not particularly care, Buchanan argued, about inequality or even deprivation per se. He argued that:

> It is not the low income of the family down the street that bothers most of us; it is the fact that the family lives in a dilapidated house and dresses its children in rags that imposes on our sensibilities. And we are willing to pay something to remove this external effect; it is relevant for behavior. (Buchanan 1968, 189)

If inequality itself did not assault voters’ “sensibilities” then “substantial transfers of general purchasing power” (Buchanan, 1968, 189) are not warranted.

When the negative income tax was proposed at the outset of the Nixon administration, Buchanan aired his skepticism to fellow members of Nixon’s Taskforce on Urban Affairs. Reacting to a draft of the Taskforce Report, Buchanan wrote that “reality suggests that the negative income tax can never substitute for other programs. For this reason, I am opposed to its enactment”. In his initial list of “policy objectives” sent to the Taskforce chairman, Edward Banfield, Buchanan outlined a new approach to means-tested welfare as an alternative to the negative income tax. He proposed that welfare programs should have “uniform or differentially reversed welfare payment standards”. ‘Differentially reversed’ referred to using welfare policy to reconfigure incentives, which for Buchanan

---

28 Memorandum, “Statement of Dissent on Particular Points in Report of the Task Force on Urban Affairs,” n.d., Correspondence Box 79, Task Force on Urban Affairs, James M. Buchanan Papers, Special Collections Research Center, Fenwick Library, Fairfax. The statement is undated but it refers to the second draft of the Taskforce Report, which is dated December 25, 1968. In a separate January 1, 1969 letter to Edward Banfield, Buchanan refers back to his dissent, so the dissent must have been written and sent at some point in the prior week.
included “subsidies to migration toward rural areas and towards suburbs” and “subsidization of unwed females who do not bear children”.  

Compared to these earlier recommendations to use welfare policy for detailed social engineering, Buchanan’s later demogrant proposal is clearly more egalitarian. But understood in the broader context of his response to the Hochman and Rodgers (1969) Pareto optimal redistribution theory, the demogrant proposal emerges out of a new analytical approach that narrowed the scope for redistribution. Ironically, Buchanan did not narrow his views by staying within the guardrails of Pareto efficiency that he previously demanded of Musgravian and Samuelsonian public finance. Instead, he introduced the external standard of political efficiency and the risk of Hobbesian conflict as a justification for the flat tax and the demogrant. This is in contrast with Buchanan and Tullock’s ([1962] 1999) foundational principle that external standards are inappropriate and only unanimity guaranteed that no one’s preferences were imposed on another.

As Buchanan’s willingness to entertain a more expansive constitutional vision narrowed, he also began to once again entertain the possibility of franchise restrictions on net transfer recipients, and particularly voters on welfare. In 1978, six years after 1972 UVA conference, Buchanan encouraged Paul Craig Roberts (a UVA alumnus and then staff assistant to Senator Orrin Hatch) to “think of the potential” of a system where “welfare recipients could not vote”. Buchanan warned Roberts that the federal government was creating “a whole voting constituency who pays no taxes” and that “this group will vote, en bloc, for continued expansion, indeed without limit”.  

These concerns are consistent with Melinda Cooper’s (2021) point about Buchanan that anxiety over an expanded franchise beginning in the 1960s was a crucial backdrop for his role in the ‘tax revolt’ of the 1970s.

V. CONCLUSION

Geoffrey Brennan (1975, 240) provides a useful distillation of the emergence of Pareto optimal redistribution out of the intellectual environment of UVA and the Urban Institute when he writes that “the significance one attaches to the possibility of Pareto-desirable redistribution depends

---

29 Buchanan to Edward Banfield, December 10, 1968, Correspondence Box 79, Task Force on Urban Affairs, Buchanan Papers.

30 Buchanan to Paul Craig Roberts, April 10, 1978, Correspondence Box 51, Folder “Correspondence Paul Craig Roberts, 1/2”, Buchanan Papers.
largely on the sort of Paretian one is”. Harold Hochman and James Rodgers were ‘the sort of Paretian’ interested in the theory of public finance and filling the gaps in Musgrave and Samuelson. Their Urban Institute colleagues were, in contrast, ‘the sort of Paretian’ interested in making actual improvements (Pareto or otherwise) in public policy and grappling with Okun’s efficiency and equity tradeoff. Buchanan, like Hochman and Rodgers, was a Paretian theorist. But for Buchanan, Pareto had to be considered at both the constitutional and policy-making stage of analysis. This opened entirely new questions and considerations.

At the 1972 conference in honor of Milton Friedman, Buchanan’s (1974a, 77) solution was, in his words, “close to that expressed by Musgrave” in Musgrave’s comment on Hochman and Rodgers (1969) because it brought in the property rights considerations that structured the “primary distribution”. Cast in Musgrave’s (1970) terminology, Buchanan’s later demogrant proposal is better thought of as a secondary redistribution that is designed in full cognizance of the primary distribution of resources and their interaction with political power dynamics. Both Buchanan’s (1974a) property in franchise arguments and his demogrant proposal were therefore attentive to Musgrave’s distinction between primary and secondary redistribution, but addressed the problem in different ways.

A final conclusion that falls out of this study is that scholars wrestle with ideas over the course of decades, and the different ideas that they wrestle with often interact in unexpected ways. Buchanan first confronted Pareto optimal redistribution as ‘the sort of Paretian’ that wrote *The Calculus of Consent*: unanimity was imperative, and the implications of voting rules could not be ignored. Following that thread, Buchanan came to the unexpectedly expansive conclusion that the franchise implied a contingent claim to the wealth of the whole community. By the 1970s Buchanan was struggling with a different set of intellectual problems. To his eyes, the barbarians were at the gate and rather than study the parameters of theoretical constitutions he turned his attention to political arrangements that would bring order to anarchy for the benefit of all (or at least most) citizens. For Buchanan, this new research agenda implied a demogrant—a universal and expansive, but at the same time much narrower approach to redistribution and public welfare programs than he had entertained in the early 1970s.
REFERENCES


Kuehn / “The Hardest of All the Problems”


Daniel Kuehn is a Principal Research Associate at the Urban Institute’s Income and Benefits Policy Center and an adjunct professor at George Washington University, both in Washington, D.C., USA. He holds a PhD in Economics from American University and has published numerous articles, book chapters, and reports on the history of economics and on employment and training programs.

Contact e-mail: <dkuehn@urban.org>
Why ‘Indirect Discrimination’ Is a Useful Legal but Not a Useful Moral Concept

KASPER LIPPERT-RASMUSSEN
Aarhus University

Abstract: A policy (practice, act, etc.) indirectly discriminates against a group, G, if, and only if: (1) it does not reflect an objectionable mental state regarding the members of G; (2) it disadvantages members of G; (3) the disadvantages are disproportionate; and (4) G is a socially salient group. I argue that indirect discrimination is not non-instrumentally morally wrong. Clearly, if it were, that would be because it harms members of G disproportionately, i.e., in virtue of features (2) and (3). Harming members of a group disproportionately does appear non-instrumentally wrong. But it is not easy to provide a plausible explanation for the kind of harm and disproportionality involved here that vindicates this initial appearance. This does not mean the concept of indirect discrimination should be jettisoned. It was originally a legal concept, and in closing I briefly suggest that in law it plays a valuable role, even if it is not a genuine moral category. Legal prohibition is an unreliable guide to what is morally wrong, but it is not supposed to be that anyway.

Keywords: discrimination, disproportionality, distributive justice, harm, indirect discrimination

JEL Classification: D63, J71

I. INTRODUCTION

Indirect discrimination is not non-instrumentally wrong, i.e., it is not the case that it is wrong independently of its causal consequences—or so I
shall argue in this article. For all I say in this article, direct discrimination might be non-instrumentally wrong and, thus, different from indirect discrimination in this respect, but I will not address the question here. Direct and indirect discrimination are generally distinguished, and generally the distinction is treated as exhaustive and exclusive: it is assumed that, on the basis of a specific feature on which the property of being discriminatory supervenes, all cases of discrimination against a particular group are either direct or indirect, but not both (Lippert-Rasmussen 2013, 40). Most theorists take the fact that the discriminator’s treatment of the discriminatee is driven in some way by, or reflects, an objectionable mental state of his or hers to be a defining feature of direct discrimination. What that mental state is—e.g., an intention to exclude (Altman 2020), feelings of animosity (Arneson 2006), false beliefs about the discriminatee’s inferior moral status (Alexander 1992), or an indifference to the interests, or failure to attend to the autonomous choices, of the discriminatees (Eidelson 2015)—and how it is thought to explain the moral wrongness of direct discrimination varies. But for present purposes the crucial thing is that indirect discrimination is defined by the absence of any such mental state.

This absence is not the only way in which indirect discrimination differs from direct discrimination. However, it is the most obvious. Given this, it is natural to ask: When a policy (or practice, act, etc.) is not associated with an objectionable mental state, what makes it discriminatory? As the US term for indirect discrimination, ‘disparate impact’, indicates, the policy’s disadvingaging of those who are discriminated appears to be

---

1 The disproportionate disadvantage for the discriminatee is not a causal consequence of indirect discrimination (see the definition of indirect discrimination below), but a defining feature of it. To say that indirect discrimination is not non-instrumentally wrong means that instances of indirect discrimination might be wrong either in virtue of their defining features (e.g., the disproportionate advantages for the discriminatees that the policy involve in the specific cases), wrong because of their consequences, or not wrong at all.

2 This formulation allows that an instance of discrimination could at one and the same time be an instance of direct discrimination against one group and an instance of indirect discrimination against another group. Setting aside issues pertaining to social salience (see below) a rule requiring employees to be taller than 175 cm could be directly discriminatory against short people and indirectly discriminatory against women. The formulation also allows that a point-based hiring system, which includes the rule ‘Deduct 10 points from applicants who are either female or not taller than 175 cm’, is both directly discriminatory against women in virtue of the rule’s first disjunct and indirectly discriminatory against women in virtue of its second disjunct.

3 Thus direct (unlike indirect) discrimination need not disadvantage discriminatees and is not tied to disproportionality.
crucial here. Hence, for the purposes of this article I define indirect discrimination as follows:

A policy (practice, act, etc.) is indirectly discriminatory against a certain group, G, if, and only if:

1. it does not reflect an objectionable mental state regarding members of G (for short, the mental state condition);
2. it disadvantages members of G (the disadvantage condition);
3. the relevant disadvantages are disproportionate (the disproportionality condition);
4. and G is a socially salient group (the social salience condition).

The following different, but very similar, characterizations of indirect discrimination have been offered: (a) it occurs when “an act impacts members of a protected group disproportionately, in comparison with its cognate groups” (Khaitan 2018, 32). (b) It pertains to “a certain group, P, if (1) it reflects no bias on the part of the discriminator against members of P on account of them being members of P or any intent to discriminate against members of P (the no-intention condition); (2) it unjustifiably disadvantages one group relative to another relevant contrasting group (the relative disadvantage condition); and (3) P can be considered as a group in the relevant sense (the group condition). All three are necessary to identify cases of indirect discrimination” (Cosette-Lefebvre 2020, 342). (c) “Where a general policy or measure has disproportionately prejudicial effects on a particular group, it is not excluded that this may be considered as discriminatory notwithstanding that it is not specifically aimed or directed at that group” (Shanaghan v. United Kingdom, App. No. 37715/97 Eur. Ct. H.R., para. 129 [2001]; quoted in Altman 2020). (d) “A practice indirectly discriminates against a person, P, on the basis of t, if P has t, P is disadvantaged by the practice, and although the practice does not explicitly single P out because of t or some related trait, u, it nevertheless disproportionately disadvantages those who have t relative to those who have not” (Moreau 2020, 17). (e) Indirect discrimination occurs where “a facially neutral measure [...] impacts more harshly on one group than on another” (Doyle 2007, 538; see also Fredman 2018; Hallde- nius 2005, 459; Khaitan 2015, 73–76; Thomsen 2015). As Doyle (2007, 537–538) points out, on the so-called purpose model of indirect discrimination a facially neutral criterion is used to exclude members of a particular group. Such cases are best thought of as cases of (concealed) direct discrimination, so I ignore them here.

In a legal context, business necessity or a suitable relation to job performance are sometimes taken to rule out indirect discrimination (Griggs v. Duke Power Co., 401 U.S. 424, 431 [1971]). Arguably, business necessity and proper job performance-relation are covered by the disproportionality condition. For a related point about the necessity condition in the context of just war theory, see Hurka (2005, 38).

Some might adopt a narrower, asymmetric view on which only dominated socially salient groups can be subjected to indirect discrimination (for an overview of anti-subordination in legal theory, see Balkin and Siegel 2003; see also Sullivan 2004). I set aside this view here. If we adopt it, the question arises as to whether a policy that is directed at a non-dominated group and is in all morally relevant respects like an indirectly discriminatory policy against a dominated group (e.g., it disproportionately harms members of the non-dominated group) is any worse, morally speaking, than the otherwise comparable indirectly discriminatory policy. ‘Otherwise comparable’ could be understood in such a way that the same disadvantage in one sense of ‘disadvantage’ imposed on a member of a dominated group and on a member of a dominating group (e.g., not being hired) typically involves very different degrees of disadvantage in the relevant sense. For
The mental state condition certainly seems to capture a feature that we are latching on to when we see a difference between direct and indirect discrimination. As I have indicated, it is also a generalization of accounts of indirect discrimination in the literature. The disadvantage condition seems incontrovertible. In a case of direct discrimination it might make sense to complain about being subjected to discrimination qua being a member of G even if, for some reason, one is not harmed by being discriminated against. But the notion that one could benefit from being indirectly discriminated against (or indeed be unaffected by indirect discrimination) makes no sense. If anything, such cases are better described as situations in which either there is no discrimination at all or, if the members of G benefit, there is indirect discrimination in favor of G.

The disproportionality condition, likewise, is plausible. As noted by Hellman (2018, 108), not just any policy that disadvantages members of G will indirectly discriminate against G (let alone be morally wrong for this or other reasons). For one thing, it may be the case that all the alternatives to the policy in question impose greater disadvantages with greater moral significance on other groups, in which case complaining about the practice on grounds of indirect discrimination would appear to be misplaced. The disproportionality condition is designed to immunize accounts of indirect discrimination, including the one given in (1) to (4), example, losing one of very few options for being hired versus losing one of very many options for being hired. On this understanding, we may be able to explain why, say, white, middle-aged, heterosexual men are much less likely to experience indirect discrimination than members of minority groups, even if there is no conceptual barrier to their having disproportionate disadvantages imposed on them.

7 I set aside here the complication of policies that do not reflect objectionable mental states but pick out all (or only) members of the group of discriminatees for disadvantageous treatment and do not pick out any members of the group of non-discriminatees for disadvantageous treatment. Plausibly such policies are classified as directly discriminatory. However, this complication makes no difference to my arguments below, so I set it aside.

8 I set aside how indirect discrimination should be defined in relation to merely possible people—an issue that might be quite relevant in relation to abortion policies, etc.

9 I do not address what the relevant measure of (dis)advantage is in this article. My arguments are neutral regarding the proper metric of (dis)advantage.

10 It might be suggested that such a complaint makes sense on the grounds that the relevant non-instrumental wrongness thesis merely says that indirect discrimination is pro tanto wrong, not wrong all things considered (see below). One rejoinder to this suggestion notes that, typically, when people assess whether the disproportionality condition is satisfied, they assume that only relatively direct effects (whatever ‘direct’ means here) count for the purposes of assessing proportionality. On that assumption, and given that remoter effects count in relation to whether something is morally wrong all things considered, we should reject the relevant suggestion.
against this problem. Clearly, that leaves us with the task of specifying when harms are disproportionate. I return to this in section IV.

Finally, the social salience condition. A lot of policies that are not considered discriminatory nevertheless impose disproportionate disadvantages on groups of people. The groups, however, are—in a sense needing to be explained—artificial: an example is the group of people who are disproportionately disadvantaged by the policy in question. The social salience condition addresses this issue. The key idea is that a group of people can be subjected to indirect discrimination only if they form a socially salient group, where a group is socially salient if, and only if, “perceived membership of it is important to the structure of social interactions across a wide range of social contexts” (Lippert-Rasmussen 2013, 30). Men and women, African Americans and European Americans, disabled and non-disabled people are socially salient groups in the US. Americans living in a postal district whose number contains two identical numerals are not. The social salience condition finds strong support in that almost all complaints about indirect discrimination concern socially salient groups, and that, typically, in cases where socially salient groups are subjected to indirect harms in a way that does not reflect objectionable mental states, indirect discrimination is detected.\(^\text{11}\)

In this article, I want to use the definition of indirect discrimination in (1) to (4) as a springboard from which to explore the concept of indirect discrimination and ask why this form of discrimination is considered morally wrong. By ‘morally wrong’ I mean (mostly without stating this explicitly) \textit{pro tanto} morally wrong, i.e., that there is something about indirect discrimination which makes it to some degree morally wrong. An act or practice can be \textit{pro tanto} morally wrong, yet not morally wrong all things considered. An instance of indirect discrimination might also have various good-making features such that all things considered it is not wrong to engage in it. Acts and practices can also be \textit{prima facie} wrong, yet not \textit{pro tanto} wrong. Thus, the fact that an act amounts to indirect discrimination might be a defeasible reason to think that it is wrong for some reason or other even if we know that this reason cannot be that it is a case of indirect discrimination. Hence, the denial that indirect discrimination is \textit{pro tanto} wrong, leaves it open to me to say that it is \textit{prima facie} wrong in the indicated way. Section II argues that if we accept the social salience condition, we must also accept that indirect discrimination

\(^{11}\) See, however Cosette-Lefebvre (2020, 351–357) for a critique of social salience-focused definitions of discrimination.
is not a distinct wrong—in other words, cases of indirect discrimination are morally no worse than cases where comparably disproportionate harm is done to a socially ‘insalient’ group of people. From here onwards I use this handy expression to refer to groups without lacking salience.) That leaves it open that indirect discrimination is a wrong, but not distinctively so. Section III explores the disadvantage condition in that light. It considers different accounts of the conditions under which indirect discrimination harms the discriminatees with a view to determining whether harming them in one of these ways might be non-instrumentally wrong. Section IV turns to the disproportionality condition, which is underspecified in the above definition. I argue that friends of the view that indirect discrimination is non-instrumentally wrong face a dilemma. Either we understand disproportionality in a non-revisionist way, or we do not. If the former, then indirect discrimination cannot be non-instrumentally wrong. If the latter, it can—at least, this seems a genuine possibility—but then, of course, we are saddled with a revisionist view of indirect discrimination.

I am not the first theorist to defend the claim that indirect discrimination is not non-instrumentally wrong. The main novelties of this article lie elsewhere. First, it offers several new arguments for this claim that are based on a relatively detailed analysis—a second novelty—of group disadvantage and the disproportionality conditions. In section V it also offers an account of why the concept of indirect discrimination is useful despite the arguments of the previous section. The view that indirect discrimination is not non-instrumentally wrong might be taken to mean that we need not be concerned about that form of discrimination. However, this inference should be questioned. In section V, I explain why, in general, there may well be good moral reasons for being concerned, from the point of view of law, about policies (practices, acts, etc.) that are not

---

12 There is another sense in which the wrong of indirect discrimination can be said to be distinct: in virtue of being different from the wrong of direct discrimination. Although this is not the distinctiveness I reject in section II, Sophia Moreau (2020) does deny that the wrong of indirect discrimination is distinctive in this sense. On her view, both direct and indirect discrimination are wrong because they involve a failure to treat the discriminatees as moral equals.

13 I defended the same view in Lippert-Rasmussen (2015) and Eidelson (2015, 39–68) defends a similar view. Also, theorists—say, libertarians—who think that discrimination is wrong only in a quite narrow range of cases take the same view as I do on whether indirect discrimination is non-instrumentally wrong (e.g., see Cavanagh 2002). Recently, Moreau (2020) has defended the claim that direct and indirect discrimination are both non-instrumentally wrong, and for the very same reason, i.e., that they involve failing to treat the discriminatees as equals.
wrong in themselves. It argues that these reasons may well be applicable to indirect discrimination. Section VI concludes.

II. SOCIAL SALIENCE

Consider a policy which is disproportionately harmful to a particular group of people who do not form a socially salient group.\(^{14}\) Compare that policy to one that is exactly alike except that the people harmed do form a socially salient group. My contention is that these two cases cannot differ, morally speaking. Why not?

The most obvious answer is that a given harm is not made smaller simply by being imposed in a particular way—in this case by being perpetrated in a way that visits the harm on members of a particular (socially insalient) group. Hence, insofar as it is differences in the magnitude of harm that make a difference to wrongness, then the social salience or otherwise of the group cannot make the two cases morally different. It is true that harm is unlikely to be the only wrong-making feature. Arguably, objectionable mental states such as animosity (see section I) are another. However, because our concern here is with indirect discrimination and relevantly similar policies, the two cases I am comparing do not differ in those terms.

It might be replied that something must have gone wrong here. Imposing a certain amount of disproportionate harm on, say, a vulnerable, socially salient minority seems to be morally worse than imposing the same amount of harm on a group of people who do not form a socially salient group and who are not similarly vulnerable. I concede that the two cases do indeed look morally different. However, this is entirely consistent with the claim that I am making here. Note, first, that a group of people can be vulnerable even if they do not form a socially salient group. Hence, if we assume that the two cases differ in terms of vulnerability, we are comparing cases of disproportionate harm to two groups of people that differ only in the fact that one group is socially salient, and the other is not.

Second, certain vulnerabilities are closely tied to membership of a socially salient group. For example, the members of a socially salient group, G, may be vulnerable to negative stereotypes about members of G in ways that it is difficult to see members of a socially insalient group being. Presumably, groups that are not socially salient typically do not attract

\(^{14}\) Recall my definition of a socially salient group in section I.
stereotyping (Beeghly 2021). But while this is a morally relevant difference, it is not a difference which by definition—as opposed to: as a general matter—overlaps with the distinction between groups that are socially salient and groups that are not. Accordingly, it cannot show that social salience is non-instrumentally morally significant, and thus it cannot show that imposing disproportionate disadvantages on a socially salient group is morally worse than imposing those disadvantages on a group which is not.¹⁵ I conclude that while indirect discrimination might be wrong, if it is, it is not distinctively wrong—by which I mean ‘wrong in a way that otherwise comparable instances not amounting to indirect discrimination are, by contrast, not wrong’. The question then is: Is indirect discrimination non-instrumentally wrong?¹⁶ My response to this question will occupy the next two sections.

III. GROUP DISADVANTAGE

The disadvantage condition says that indirectly discriminatory policies disadvantage members of the groups being subjected to discrimination. This condition is rarely specified in the respects I want to consider, but it can be understood in several ways. Consider the classic court case involving indirect discrimination, Griggs v. Duke Power Co. (401 U.S. 424 [1971]).¹⁷ Here, is the disadvantage condition satisfied if the company’s

¹⁵ Something similar is true of expressive harms. While it might be true that, virtually always, only members of socially salient groups are subjected to expressive harms qua their group membership, in principle members of insalient groups could suffer expressive harms too. For further arguments against it being non-instrumentally morally relevant whether a group of people disadvantaged by a certain policy form a socially salient group, see Thomsen (2013).

¹⁶ On the desert-accommodating, prioritarian account of the wrongness of discrimination that I proposed in Lippert-Rasmussen (2013), it is both the case that the wrong of indirect discrimination is not distinct from the wrong of direct discrimination and that the wrong of discrimination, whether direct or indirect, is not distinct from otherwise comparable, but non-discriminatory, policies etc. Non-distinct moral wrongs can, in my sense, be serious moral wrongs. Note also that while section II defeats some suggestions as to what might make indirect discrimination distinctively wrong, it does not demonstrate that there cannot be any feature of indirect discrimination which makes it distinctively, non-instrumentally wrong.

¹⁷ This 1971 US Supreme Court verdict implied that a hiring policy can be illegal when it is “fair in form, but discriminatory in operation” (Griggs v. Duke Power Co., 401 U.S. 424, 431 [1971]), i.e., indirectly discriminatory. The employer, Duke Power, had “instituted requirements of high school education and satisfactory scores in an aptitude test as a condition of employment or transfer. The same test was applied to all candidates, but because black applicants had long received education in segregated schools, both requirements operated to disqualify black applicants at a substantially higher rate than whites” (Fredman 2011, 178). Since the requirements went beyond what was needed to ensure satisfactory levels of performance, the company was ordered by the court to abolish them.
seniority policy disadvantages its own African American employees? Or is it satisfied if that policy (or the set of similar policies operated by several companies including Duke Power) disadvantaged African Americans in general? It probably resulted in both, even if its effects on the Duke Power employees were much more significant than its largely marginal effects on African Americans generally. From our point of view the crucial question is about which of following four views articulates the correct understanding of the disadvantage condition (for convenience, I consider discriminatory policy from this point onwards):

A policy satisfies the disadvantage condition if, and only if:

Local View: it disadvantages a relevant subset of G, e.g., employees in the relevant company.

Global View: it disadvantages members of G in general, e.g., members of G on average or each individual member of G.

Broad View: it satisfies either the local or the global view.

Restrictive View: it satisfies both the local and the global view.

I think that insofar as indirect discrimination is non-instrumentally wrong both local and global disadvantages must matter. In my view, an act or practice can qualify as wrongful, indirect discrimination either because of the disadvantages it imposes on a subset of the relevant group members or because of the disadvantages it imposes on group members in general even if it does not involve both forms of disadvantage. If so, the Restrictive View is too restrictive. Consider the following case, designed to test the supposition that global effects are irrelevant to whether something amounts to discrimination:

Hypothetical Griggs v. Duke Power Co.: This case is like the actual one except that while Duke Power's promotion rules disadvantage its African American employees, for some weird reason this benefits African Americans in general.

It is too restrictive because, presumably, a policy may be indirectly discriminatory if it disproportionately disadvantages all the relevant company's employees even if it does not disadvantage members of the relevant minority in general.

It is standard in many philosophical disciplines to appeal to our intuitions about hypothetical examples like the present one (in fact, often examples much more outlandish than this one are presented). However, some theorists are sceptical of whether we can learn anything from the intuitions we have about hypothetical cases (see also Moreau
Plausibly, for some sufficiently small size of the relevant local harms and some sufficiently large size of the global benefits for other members of the relevant group, the connected policy would not be *pro tanto* wrong even if it were indirectly discriminatory. Granting this, and granting also that indirect discrimination is non-instrumentally wrong, it follows that the disadvantage condition does not apply only to harms to the relevant subset of the group of people being subjected to indirect discrimination. Perhaps in many cases indirect discrimination only disadvantages members of the relevant group locally, and we may wish to say that this suffices to satisfy the disadvantage condition. However, that does not mean that global disadvantages are irrelevant. It also follows that on a non-revisionist understanding of it, indirect discrimination is not non-instrumentally wrong, since the local disadvantages could be substantially outweighed, morally, by the global advantages to the group to which the discriminatees belong.

Whether we accept the Broad View, as I have just suggested we should, or instead accept one of the three alternative interpretations of the disadvantage condition, there is a further question about what the disadvantage condition says, namely: What exactly is it to disadvantage the relevant (subset of the) group in question?

There are simple cases where this question is not pressing, because *all* members of the relevant group are disadvantaged by the indirectly discriminatory policy. However, real-life cases of indirect discrimination—especially if the disadvantages condition applies globally—are rarely simple. Typically, some members of a group subjected to indirect discrimination are disadvantaged and others are not affected—and some might even benefit all things considered. Hence, we need to ask when a group is disadvantaged all things considered. The most obvious answer is:

*The Average View*: Members of a group are disadvantaged by a certain policy if, and only if, the policy disadvantages members of the group on average.

While the Average View is a natural view to take, and although it has intuitive results in a range of cases, it cannot be right if indirect discrimination is non-instrumentally wrong—at least, not if we adopt a widely

---

2020, 29). I do not have the space to discuss this methodological issue. For a view which is moderately sympathetic to the use of thought experiments in applied ethics, see Walsh (2011).
accepted account of the aggregation of harms and benefits. Consider a
case where the great majority of group members benefit marginally from
a policy while a minority suffer very significant harm from it—so signifi-
cant, in fact, that people who oppose unrestricted aggregation would say
that it is morally impermissible to impose that very significant loss on a
few people to benefit a much greater number of people even though the
aggregate benefits are much greater than the aggregate harms (cf. Scanlon
1998, 229–241; see also Voorhoeve 2014). Here we might be inclined to
say that the policy mistakenly alleged to be indirectly discriminatory is
wrong because it disproportionately harms members of a certain group
even if members benefit on average and even if most members of the
group benefit. However, we might think instead that if indirect discrim-
inination names a particular moral wrong, it is better to classify it as a case
of indirect discrimination. With this attitude, we might adopt:

*The Anti-Unrestricted Aggregation View*: Members of a group are dis-
advantaged by a policy if: (i) some of its members are severely disad-
avantaged by the policy; (ii) the aggregate disadvantages imposed on
these members is smaller than the aggregate benefits conferred on
other members of the group; (iii) but the individual benefits conferred
on these members are normatively insignificant relative to the harms
to the former.

Suppose we are anti-aggregationists. That is, we believe that there are
some trivial benefits which, by comparison with some significant harms,
are such that, however large the number of people receiving the former
and however small the number of people suffering the latter, it is morally
impermissible to impose significant harms of that type on the smaller
number of people even if that brings about a (much) greater sum of

---

20 This claim does not rest on any assumptions about how well off the recipients of the
marginal benefits and the recipients of the significant harms are—either absolutely or
relatively—before being benefited or harmed in the relevant ways. The claim is con-
sistent with the view that such assumptions make a difference to whether a benefit or
harm of a certain magnitude (i.e., a certain amount of prudential value) counts as mar-
ginal or significant from a moral point of view.

21 This case also enables us to see that another intuitively appealing view—the Numbers
View, which says that members of a group are disproportionately disadvantaged only if
a (super)majority of the group’s members are disadvantaged—is false. Thomsen (2015,
321) describes a variant of the Numbers View: a policy disadvantages P-persons if “a
greater proportion of P-persons than of P-persons are negatively affected by the treat-
ment, or [...] P-persons are more severely affected than P-persons by the treatment”.
Assuming indirect discrimination is wrong non-instrumentally, as Thomsen (2015, 322)
in effect acknowledges, this view is problematic for a reason like that defeating the Num-
bers View.
benefits. If that is the case, and if indirect discrimination is non-instrumentally wrong, we must reject the Average View and accept the Anti-Unrestricted Aggregation View. At any rate, not doing so would render the fact that a certain practice is not indirectly discriminatory less significant, since the absence of indirect discrimination would then be compatible with the impermissible imposition of severe harms on a small subset of group members. Even worse, there could be cases of indirect discrimination which are not non-instrumentally wrong, because the only alternative involves the imposition of severe, but aggregately smaller, harm on a tiny subset of group members.

Even setting aside the issue of unrestricted aggregation, a further reason to reject the Average View is that it is insensitive to the distribution of harms and benefits across members of a group. If that distribution is morally significant, it will be a factor in when we want to say that the disadvantage condition is satisfied. In the interest of simplicity, let us focus on a prioritarianism—the view that benefiting “people matters more the worse off these people are” (Parfit 1991,19). On this view, two policies imposing the same amount of overall harm on the same number of members of a particular group could differ in that one is indirectly discriminatory, and the other is not. In one case most of the harm is done to worse off members of the group, whereas in the other most of the harm is done to better off members. Assuming the policy also delivers certain benefits to members of the groups such that, on the Average View, none of them counts as disadvantaging members of the group, we might think that such a case is a reason to reject the Average View, since, given its terrible distributive profile, the policy in question here does, in the normatively relevant sense, disadvantage members of the group. Instead, we might adopt:

The Distribution-Sensitive View: Members of a group are disadvantaged by a certain policy if, and only if: the policy disadvantages members of the group on average; and the relevant disadvantages are weighted according to how badly off the recipient of the disadvantage is—the worse off she is, the greater weight the disadvantage has.

Again, it seems we cannot consistently accept that distributive justice involves distribution-sensitivity, the Average View, and the view that indirect discrimination is non-instrumentally wrong. If we accept the Average View, we might have to classify a policy as indirectly discriminatory and yet accept that it is not even pro tanto wrong because any alternative
would involve a morally worse distribution of harms, e.g., one where harms of a similar size fall mostly on people who are already worse off.\footnote{I assume that the mere fact that there is an individual who is worse off under a certain distribution than she would be under some alternative distribution is not enough to make it \textit{pro tanto} wrong (as opposed to being in one respect worse). An option I am drawn to here is that it is only if the individual is worse off under a particular distribution relative to the morally required distribution that her being worse off is \textit{pro tanto} wrong.} Equally, we would have to recognize the possibility of policies involving a bad distributive profile which we could not classify as indirectly discriminatory because on average the group members benefit. Yet we would consider these policies unjust. Again, this would render it less significant that a certain policy is not indirectly discriminatory. After all, that sought-after status would then be entirely consistent with the severe harms imposed on worse off members of the group by the policy rendering it impermissible. Hence, the consideration regarding non-aggregation and distribution suggests that on a non-revisionistic understanding of the group disadvantage condition indirect discrimination is not non-instrumentally wrong.

Finally, the views discussed so far imply that we can assess whether something amounts to indirect discrimination without knowing anything about the causal history of the disadvantages in question. Arguably, this is not the way standard cases of indirect discrimination have generally been understood. Take \textit{Griggs v. Duke Power Co.} again. There, the company’s promotion rules harmed African Americans as a result of historical (direct) discrimination against them, since that meant that, as the rules were applied, no African Americans had the seniority needed for promotion. The promotion rules compounded previous unjust harms. Had the case not involved this feature—had it been untrue, in other words, that the distribution of seniority when the rules were applied was itself expressive of discrimination—then, arguably, the disadvantage condition would not have been satisfied.\footnote{This is consistent with the claim that almost any disadvantage imposed by a policy on standard victims of discrimination compounds an injustice. My case here simply rests on the modest contention that it is \textit{possible} for the disadvantage not to be a compounding of a prior unjust disadvantage. For a real-life case which arguably did not involve compounded injustice, yet was seen by courts as a case of indirect discrimination, see Hellman (2018, 121).} The view underpinning this asymmetric assessment might be:

\textit{The Anti-Compounding View:} Members of a group are disadvantaged by a certain policy in the relevant sense if, and only if, the policy...
disadvantages members of the group, and those disadvantages are causally dependent on prior unjust discrimination against this group.\textsuperscript{24}

On this view, a policy does not indirectly discriminate against a group it disadvantages if the disadvantages are causally independent of prior unjust discrimination against the group.\textsuperscript{25} However, indirect discrimination might take place if the disadvantages in question are so dependent. If we think that indirect discrimination is wrong when it compounds prior injustice (Hellman 2018, 106), and if indirect discrimination is non-instrumentally wrong, it seems to follow that the Anti-Compounding View of disadvantage is crucial to a correct understanding of the disadvantage condition. At least, if disadvantages are understood in terms of anti-compounding—something which Hellman (2018, 121) acknowledges is somewhat revisionist relative to standard legal notions of indirect discrimination—it seems to follow that indirect discrimination is non-instrumentally wrong. If, however, indirect discrimination is wrong when it compounds prior injustice, and if we reject the Anti-Compounding View of disadvantage, it follows that indirect discrimination is not non-instrumentally wrong.\textsuperscript{26} This follows because, given these two assumptions, there will be cases of indirect discrimination that do not compound injustice, since the disadvantage condition can be satisfied in the absence of any compounding of injustice.

To sum up, the disadvantage condition is much more complex than it looks. Specifically, if indirect discrimination is non-instrumentally morally wrong, we must, in a revisionist way, take that condition to pertain to both local and global disadvantages. We then face a question about how to weigh, say, local disadvantages against global benefits and vice versa. Also, it seems we must reject the Average View. Finally, we might even have to take a Restrictive View of what counts as a disadvantage in the

\textsuperscript{24} The Anti-Compounding View raises the same issues of interpretation that the Average, the Anti-Unrestricted Aggregation, and the Distribution-Sensitive Views all address. The causal dependency condition is related to a similar, but distinct, causal dependency condition proposed in Lippert-Rasmussen’s (2013, 71) definition of indirect discrimination. The latter refers to past or present direct discrimination, while the present condition refers to prior unjust discrimination, thus including unjust indirect discrimination.

\textsuperscript{25} A slightly broader view includes injustices other than unjust discrimination.

\textsuperscript{26} The present point does not require us to accept Hellman’s (2018) view that indirect discrimination is wrong because it compounds injustice, and here I am simply using that view as a pointer to a certain interpretation of the group disadvantage condition. In passing, however, I should mention that I am not persuaded by Hellman’s account for reasons explained in Eidelson (2021) and Lippert-Rasmussen (forthcoming); for a reply to Eidelson, see Hellman (2021).
relevant sense—that is, insist that it is only disadvantages compounding prior injustice that so count. I have not shown that there is no possible interpretation of ‘disadvantage’ on which indirect discrimination can be non-instrumentally wrong, but I have shown, I hope, that what such a view would look like is significantly more complicated than is normally assumed. I also hope to have shown that adopting a view of disadvantage on which it is possible to claim that indirect discrimination is non-instrumentally wrong requires one to take a distinctly revisionist view of several dimensions of the cases that are to be counted as genuine cases of indirect discrimination. The dilemma this generates has not, as far as I am aware, received attention in the literature so far.

IV. DISPROPORTIONALITY

Suppose we have come up with a convincing answer to the issues discussed in the previous section. That still leaves us with the challenge of explaining the disproportionality condition. This challenge has two aspects. First, if a disadvantage is disproportionate, it is disproportionate relative to something else. But what, exactly, is the thing we are to compare the disadvantage to the indirect discriminatees with in order to determine whether its imposition is disproportionate? Call this the issue of the proper comparanda. Second, once we have fixed what the proper comparanda are, we need to determine what it is for them to be related in a way exhibiting disproportionality. For example, is it sufficient for the disadvantage to indirect discriminatees to be greater to some degree than the corresponding benefits to others or must it be significantly greater? Call this the issue of quantity. In this section, I address both issues with a view to determining whether indirect discrimination can be a non-instrumentally wrong.

Let me start with the issue of the proper comparanda, drawing on (Lippert-Rasmussen 2013). The two most interesting views here are:

---

27 In the previous section, I explored issues about unrestricted aggregation, and distribution, and whether the relevant disadvantage compounds a prior injustice. Similar issues come up in relation to the relevant positive comparanda, as it were. For example, do the benefits to others fall on people who are very badly off, or on people who are very well off? Suppose that a policy harms well-off women significantly but benefits badly off men slightly. On a distribution-sensitive view of proportionality, this policy might not satisfy a distribution-sensitive version of the proportionality condition. For simplicity, I set these matters aside in this section and assume that we should simply compare the size of the disadvantages to indirect discriminatees with the size of the benefits to others.
The Group Comparison View: A group, G1, is disproportionately disadvantaged by a policy if, and only if, the inequality between G1 and G2 (a group with which it is to be compared) is greater with the policy than it would be in some relevant alternative situation without the policy.28

The Advantages Comparison View: A group, G1, is disproportionately disadvantaged by a policy if, and only if, the gap between the advantages it would enjoy under that policy relative to the greater advantages that G1 would enjoy under some relevant alternative policy is disproportionately relative to the gap between the advantages some group, G2, with which it is to be compared would enjoy under that policy relative to the smaller advantages G2 would enjoy in some relevant alternative situation without it (Lippert-Rasmussen 2015).29

To see how the two views differ, consider a situation where we must choose between two policies, Equal and More. Both men and women will be best off under More, which however involves inequality in men’s favour. Under Equal policy there will be strict equality between men and women, but both groups will be worse off than they would have been under More. Here, the Group Comparison View implies that women are disproportionately disadvantaged by More—the inequality between men and women is greater under More than it is under Equal. The Advantages Comparison View need not have this implication, because neither men nor women are disadvantaged under More in comparison with how well-off they would be under Equal.

On the whole, those alleging indirect discrimination either do not distinguish between Group Comparison and Advantages Comparison. If they do, they are likely to assume that, in practice, a policy that is disproportionate in one sense will be disproportionate in the other as well.

28 Khaitan (2018), Cosette-Lefebvre (2020), and Doyle’s (2007) definitions of indirect discrimination (see note 4) all seem to imply the Group Comparison View.

29 A more formal definition of the two views is the following: Let \( A(g, p) \) be the advantage that group \( g \) enjoys under the policy \( p \). A group, \( G_1 \), is disproportionately disadvantaged (in comparison with another group, \( G_2 \)) by a policy \( a \) (as compared with an alternative policy \( \beta \)) if, and only if: The Group Comparison View: \( A(G_2, a) - A(G_1, a) > |A(G_1, \beta) - A(G_2, \beta)| \). The Advantages Comparison View: \( A(G_2, \beta) - A(G_1, \alpha) >* |A(G_2, a) - A(G_2, \beta)| \) where ‘>∗’ means not just ‘greater than’ but ‘disproportionately greater than’. I owe this definition to an anonymous reviewer. Note also that because the definiendum is a policy that disproportionately disadvantages a certain group, it does not render the definition of the Advantages Comparison View problematically circular that the definiens refers to a disproportionate gap between differences in outcome for one group relative to the differences in outcome for another group. In short, different notions of disproportionality appear in the definiendum and the definiens. I thank the editors for pointing out the need to clarify this matter.
However, as More and Equal show, the two views can be pried apart. Accordingly, to say whether indirect discrimination is non-instrumentally wrong, we need to consider both explications of the comparanda of disproportionality. There are troubles ahead whichever view we side with.

Take, first, the Group Comparison View. On this view, unless you subscribe to distributive egalitarianism in a group-focused version, you cannot regard indirect discrimination as non-instrumentally wrong. But many of those who write about justice reject distributive egalitarianism in the light of the so-called levelling down objection (or for other reasons, e.g., because they think that only inequalities between individuals have moral significance). That objection asks us to consider a situation of equality in which no one is better off in any respect than they would have been in an alternative situation without equality, and then invites us to agree that the equal situation is in no way better than the latter unequal situation (Parfit 1991). People who are impressed with the levelling down objection yet started out with egalitarian sympathies typically adopt prioritarianism. But in the prioritarian perspective, a policy that is indirectly discriminatory, if the disproportionality condition is interpreted in terms of the Group Comparison View, might be one that brings about the best outcome and thus is not non-instrumentally wrong.30

The Advantages Comparison View avoids this problem, but it creates another: the problem of revisionism. For it implies that some cases that most of us would consider to be clear cases of indirect discrimination do not satisfy the disproportionality condition. Thus consider Khaitan’s (2018, 31) description of indirect discrimination as “an apparently neutral practice or policy which puts members of a protected group (say, women) at a disproportionate disadvantage compared with members of a cognate group (say, men)”. The idea here, and elsewhere when people see indirect discrimination, is clearly that, roughly, we can simply compare the proportion of, say, male to female employees to determine whether indirect discrimination has taken place.31 However, on the Advantages Comparison View it is possible for a policy to result in, say, a greater proportion of male (or female) employees even if it is not indirectly discriminatory.

30 Even on a strict egalitarian view, disproportionality as presently understood need not be a non-instrumentally wrong-making feature. This is so because the currency of the disadvantages involved in the relevantly indirectly discriminatory rule (e.g., proportion of women employees) typically is not the currency of egalitarian justice (e.g., access to advantage [Cohen 1989]).

31 “Roughly” because the differential representation must somehow result from the policies (practices, acts, etc.) of the indirectly discriminating agent and be disproportionate relative to the benefit obtained through the policy (Khaitan 2018, 41).
On this view, indirect discrimination might be non-instrumentally wrong even if we reject distributive egalitarianism, but only if we think differently about the conditions under which a policy is indirectly discriminatory and re-classify various seemingly clear-cut cases of indirect discrimination as cases that are free of that form of discrimination.

Turning to the second issue, about quantity, there are two ways to understand the disproportionality condition here: in a moralized or non-moralized way. On the first, by definition, whether a certain disadvantage is disproportionate entails something about its moral qualities, whereas on a non-moralized understanding this is not the case. On what is probably the simplest moralized analysis, a disadvantage is morally disproportionate relative to the corresponding benefits, if, in the light of the relationship between the two, it is morally unjustified to impose it. On the moralized interpretation, if a disadvantage is disproportionate, it is morally wrong by definition to impose it on the victims of indirect discrimination. On a very straightforward non-moralized analysis, a disadvantage to the victims of indirect discrimination is disproportionate if the harm that falls on them is greater than the advantages the discrimination gives to others. Here, though this might be the case, it does not follow from the fact that a disadvantage is disproportionate that it is morally unjustified to impose it. Where the disadvantage is unjustified, that is so not because indirect discrimination has occurred, but because some independent moral principle has been infringed.

Neither understanding of proportionality is unproblematic for our purposes. If we embrace the moralized conception, we cannot say that a policy is made morally wrong by being indirectly discriminatory. Rather, one reason that policy is indirect discrimination is that it is morally wrong.32 However, many of those who write about discrimination think that the fact that something is indirect discrimination makes it wrong. Hence, they need to opt for the non-moralized view.

On the non-moralized view, however, discrimination simply falls out of the picture in the following sense: it is the fact that indirect discrimination involves the imposition of disadvantages on one group of people—where these are in a suitable descriptive sense disproportionate relative

---

32 By way of analogy, consider murder. Murder is a moralized concept in that, roughly, murder is a morally wrongful type of killing. But even if something’s being murder entails that it is morally wrongful, what makes it morally wrongful cannot be that it is murder. Rather, part of what makes the action in question murder is that it is morally wrongful. For a helpful discussion of the distinction between what makes something discrimination and what makes it wrong, see Ishida (2021).
to the benefits to others—that makes it wrong. However, as I argued in section II, there is no reason to think the wrongness of such an imposition is tied narrowly to indirect discrimination, since, among other things, it could also arise in connection with harms to members of a socially insalient group. Moreover, it is unclear that this kind of disadvantage-imposition is wrong in general. Indirect discriminators typically adopt all sorts of policies—e.g., about where to build facilities, what sort of research to invest in, which goods to produce, and so on—that benefit some people and harm others (Hellman 2018, 108). Like decisions about recruitment and promotion, all these policies could be assessed in terms of disproportionality. However, few of us believe that they should be so assessed, and certainly not according to the standards of proportionality used in typical indirect discrimination cases. But if indirect discrimination is morally wrong because of its disproportionality, so must be decisions about where to build facilities, what sort of research to invest in, and which goods to produce.

Can we say that this shows many more decisions are wrong because they are indirectly discriminatory than we suppose? That might be a sensible response for those who think indirect discrimination names a particular moral wrong. After all, it is hard to see why, say, hiring policies could be pro tanto morally wrong in virtue of their disparate effects on minorities, whereas investment policies with similar disparate effects are not. Ultimately, however, many will be reluctant to go down this particular route given the radical expansion of policies which, potentially at least, could then be shown to amount to wrongful indirect discrimination. They will instead prefer to respond to the present line of argument by rejecting the view that indirect discrimination is non-instrumentally wrong and then explaining why—for pragmatic reasons connected, for example, with the successful operation of the law—we nevertheless have reason to treat indirectly discriminatory hiring decisions differently from many other decisions resulting in disproportionate disadvantage being imposed on others.

---

33 Judging by their behaviour, most people think it is permissible in daily life to make decisions about how to spend one’s money which are strongly disproportionate because they involve tiny benefits for them (dining out at a fancy restaurant) and not preventing much greater harms to others (by providing them with medicine through a donation to Oxfam).

34 Indeed, it is an exception that decisions are considered in the light of whether they involve disproportionate harms to different socially salient groups. Probably, in relation to many decisions, many of us would say that it is the relevant agent’s prerogative to select an option that is disproportionately harmful according to the standards of proportionality used in typical indirect discrimination cases.
certain groups. Something along the latter lines is what I shall briefly pur-
sue in the next section.

V. WHAT IS THE FUNCTION OF THE CONCEPT ‘INDIRECT DISCRIMINATION’?
Let me stress again that the alternative to thinking that indirect discrimi-
nation is non-instrumentally wrong is not to think that many of the poli-
cies that are rightly considered indirectly discriminatory should be per-
mitted by law. For we might think that indirect discrimination is not non-
instrumentally wrong, but that nevertheless it is a good thing that law
forbids indirect discrimination. In fact, there are two good reasons for
thinking this is indeed so. The first is that indirect discrimination law is
an instrument we can use to secure certain socially desirable ends, and
no clearly better alternative is available. The second is that if we treat
indirect discrimination in this way, we can set aside many of the puzzles
above as irrelevant, or as relevant but no threat to the desirability of in-
direct discrimination law.

Let me start with how indirect discrimination law is a useful instru-
ment even if the discriminatory policies it outlaws are not non-instrumen-
tally morally wrong. Many Americans and foreign observers believed that
the ban of direct discrimination introduced in the era of the civil rights
movements would result in greater racial equality and integration. This
did not happen. One reason is the difficulty proving that an employer has
discriminatory intent. A related reason is the limited incentive to avoid
discrimination that a discrimination law only covering direct discrimina-
tion creates. Indirect discrimination law avoids this problem (Collins and
Khaitan 2018, 25–26). Once such a law is in place, just about any signifi-
cant underrepresentation of a protected group will oblige the employer
to prove that the skewed staffing is not due to discrimination. Shifting
the burden of proof from complainants to employers in this way has no
doubt considerably strengthened the employer’s incentive to recruit in
ways that promote equality and racial integration.35 No instrument is

35 An anonymous reviewer suggested that the notion of indirect discrimination can play
a similar useful role in applied ethics in that significant underrepresentation of a group
in a way that amounts to indirect discrimination against the group is a prima facie rea-
son to think that the policies producing such underrepresentation are morally wrong.
Hence, the concept of indirect discrimination is a useful heuristic, as it were, in applied
ethics because showing that certain policies are indirectly discriminatory shifts the bur-
den of proof from those who think the policies are unjust to those who think they are
not. I accept this point. When I submit that indirect discrimination is not a useful moral
concept I have in mind the purposes of normative theory—to develop an account of
perfect, of course, and no doubt indirect discrimination law has negative side-effects. However, the crucial point here is that there is an important reason in its favor—its contribution to racial equality and integration—even if indirect discrimination is not non-instrumentally morally wrong.

So much for the outcome-based case for indirect discrimination law. How does viewing indirect discrimination law as a useful instrument rather than a regulator of policies, practices and acts that are inherently wrong help with the problems described in the three previous sections? I shall take the three conditions of indirect discrimination that I have examined above in turn.

Consider, first, the social salience condition. I argued that there is no moral difference non-instrumentally between a policy that disproportionately disadvantages a socially salient group and an otherwise similar policy that disproportionately disadvantages a socially insalient group. If we regard indirect discrimination law as a useful instrument mitigating specific injustices, this observation is neither here nor there. For, first, the fact that two wrongs are morally similar does not show that an instrument that serves well in preventing one of them will effectively prevent the other. Second, even if there is no non-instrumental moral difference between a policy that disadvantages a socially salient group and an otherwise similar policy that disadvantages a group lacking that salience, it may still be the case that there are many more policies of the former kind than the latter. Perhaps, for example, it takes less for a disadvantage to be disproportionate when it affects underprivileged groups than it does when it falls on privileged people—and if so, it may well be sensible to prohibit indirect discrimination against certain ‘protected groups’ without prohibiting any imposition of disproportionate disadvantage on any group (see note 6).

Next, I argued that the disadvantage condition needs to be specified in several key dimensions. In particular, we need to know whether the disadvantage is local or global, and whether it depends on (a relevant subset of) group averages or something more complicated. In the context of law, these issues appear much less pressing, because, as indicated in section III, in real life, and for the protected groups typically covered by indirect discrimination law, it is likely that the relevant groups will be disadvantaged whichever of the specifications we embrace. No doubt, there will be exceptions, and in those cases discrimination law will not serve
our moral aims optimally. However, a suboptimal instrument can be the best instrument available.

Consider, finally, the disproportionality condition. In section IV, I discussed the comparanda and quantity of disproportionality. If we treat indirect discrimination law as a useful legal tool, we might respond to the first issue in the way I just responded to the issue of specifying the relevant sense of disadvantage. Thus, we might say that, by and large, when a protected group is disadvantaged by a policy on the Group Comparison View, it will be disadvantaged by that policy on the Advantages Comparison View, and vice versa, and that accordingly we can regard indirect discrimination law as a useful instrument whichever view we embrace. Responding to the issue of quantity, we can simply observe that indirect discrimination law typically errs on the side of caution: arguably, whichever plausible view of proportionality we adopt, indirect discrimination law typically outlaws policies only if they are significantly disproportionate and is tolerant of policies if they are less significantly so even when—perhaps, ideally—law ought not to be tolerant of policies that involve moderately disproportionate disadvantages being imposed on certain groups. Accordingly, if our aim is to defend existing indirect discrimination law (not to embrace an expansion of the scope of existing indirect discrimination law) failure to provide a satisfactory answer to the question of quantity is not a big problem.36

The notion of indirect discrimination, then, may well be useful in a legal context, because anti-indirect discrimination law could promote or otherwise serve important moral values. I have sketched an argument to this effect, but not actually given it. In addition, I have briefly shown why, if we think about indirect discrimination in this way, the objections to the view that indirect discrimination, understood in a non-revisionist way, is non-instrumentally wrong considered in other parts of the paper (or suitably adjusted variants of them) need not worry us insofar as we adopt the perspective on indirect discrimination I have explored and supported.

VI. CONCLUSION
I have argued that, on reflection, we can see that there are real challenges in understanding indirect discrimination as something that it is non-instrumentally wrong. More briefly, and pointing in the direction I think

36 This is not to say that the issue of quantity never gives rise to problems for the proposed law-centric view of indirect discrimination. For one thing, I have not shown that anti-indirect discrimination law is never overinclusive.
that this leads, I have tried to show that it makes good sense to regard indirect discrimination as a useful legal concept, nevertheless. Legal moralists think that if something is morally wrong, we have reason to make it unlawful. What we might call ‘moral legalists’ are inclined to think that if something is unlawful it is pro tanto morally wrong (and would remain so even if it were not prohibited by law). I believe both outlooks are false, and in section V I have offered some reasons for thinking that the legal status of indirect discrimination is a poor guide to the moral status of indirect discrimination.

While I am not the first theorist defending this view about the moral status of indirect discrimination (e.g., see Eidelson 2015, 68), many writers have taken a different view (e.g., Moreau 2020, 187; Moreau 2018). Hence, it might be useful to end this article by highlighting three issues that are not part of the disagreement between those—myself included—who think that indirect discrimination is not non-instrumentally morally wrong and those who do. First, everything I have said above is consistent with saying that direct discrimination is non-instrumentally morally wrong. Second, the same goes for the view that most (apparent) instances of indirect discrimination are morally objectionable. For example, it may be that most such instances are really cases of direct discrimination (Eidelson 2015, 6), or it may be that either on their own or together with other similar policies the relevant instances result, or have resulted, in systematic, unjust disadvantages for certain groups. Finally, and partly in light of the previous point, nothing in this article motivates a strongly revisionist stance regarding indirect discrimination law.

REFERENCES


For an overview of a range of views about the role of indirect discrimination law, see (Collins and Khaitan 2018, 25–29).

I do not accept that it is (Lippert-Rasmussen 2013). However, my reasons for this are independent of the case I have made here in connection with indirect discrimination.


**Kasper Lippert-Rasmussen** is Professor of philosophy at the department of Political Science at Aarhus University. He works in political philosophy, and has conducted extensive research on discrimination, egalitarianism, and fairness. Presently, he is the director of the Centre for Experimental-Philosophical Studies of Discrimination funded by the Danish National Research Foundation.
Contact e-mail: <lippert@ps.au.dk>
The Philosophy and Economics of Measuring Discrimination and Inequality: A Roundtable Discussion

KASPER LIPPERT-RASMUSSEN (Copenhagen, 1964) is professor of philosophy at the department of Political Science at Aarhus University. He works in political philosophy, and has conducted extensive research on discrimination, egalitarianism, and fairness. Presently, he is the director of the Centre for Experimental-Philosophical Studies of Discrimination funded by the Danish National Research Foundation.

XAVIER RAMOS (Barcelona, 1968) is professor of economics at Universitat Autònoma de Barcelona. He is a research fellow of IZA and founding member of QUALITAS (Economics of Inequality and Poverty Analysis). His research focuses on economic inequality and poverty. He is currently working on equality of opportunity, intergenerational mobility, and public policy evaluation.

DIRK VAN DE GAER (Aarschot, 1965) is professor of microeconomics and public economics at the department of Economics at Ghent University. His research focuses on inequality of opportunity, dynamic aspects of income inequality and income mobility, inheritance and wealth inequality, as well as optimal taxation.

The Erasmus Journal for Philosophy and Economics (EJPE) interviewed Kasper Lippert-Rasmussen, Xavier Ramos, and Dirk Van de gaer as part of a roundtable on the philosophy and economics of measuring discrimination and inequality, hosted on May 13, 2022. The interview is divided into four sections, which cover: the concepts of discrimination and inequality (section I); the current state of the literature on measuring discrimination and inequality (section II); the relevance of measuring discrimination and inequality for policymaking (section III); and the future of measuring discrimination and inequality (section IV).

EJPE’s Note: This interview was conducted by Lennart Ackermans and Annalisa Costella, editors at the EJPE. EJPE acknowledges generous funding by the Erasmus Institute for Philosophy and Economics (EIPE) to host the roundtable.

Footnotes throughout the interview contain editorial annotations directing readers to the relevant literature discussed in the text and definitions of technical concepts used by the interviewees.
I. The Concepts of Discrimination and Inequality

EJPE: Let us begin with some preliminary questions to define the terms. Let’s start with discrimination. Some authors argue that discrimination is best understood as a moralized concept, as something that’s inherently wrongful. Others argue that discrimination is best captured as a non-moralized concept, according to which discrimination amounts to differential treatment, but is not necessarily wrongful. What do you think about these two approaches?

Kasper Lippert-Rasmussen: I think my general approach to this question is that discrimination is a term that’s used in different senses. Some people on some occasions use it in a moralized sense and on other occasions people use it in a non-moralized sense. So, I guess that the main thing is to make clear what sense of discrimination one has in mind. In some of my own work, I’ve adopted a non-moralized definition of discrimination. One advantage of doing so—for the specific purposes I have in that book—would be that, in terms of how discrimination is used in ordinary language, there are at least a number of practices of which people tend to agree they are discriminatory; and yet, they have different views (or perhaps no views) on whether the relevant differential treatment is wrongful.

One example that I discuss is the case of age discrimination, in cases of setting priorities in relation to the distribution of scarce life-saving organs. Many people would say that it is age discrimination when younger patients are prioritized over older patients, but only some of these people will say that this is necessarily unjust. Some of them might even think that it is actually required by justice to give priority to younger patients, in light of the fact that younger patients typically have enjoyed fewer good life shares than others. I think there are other examples, such as various forms of statistical discrimination, where people tend to agree that there is discrimination, but do not necessarily agree on whether it is unjust or whether it is all-things-considered morally justified. One advantage of adopting a non-moralized version of discrimination is that people who disagree about the moral qualities of various kinds of differential treatment can, so to speak, describe their topic as one of discrimination.

Could it be a disadvantage, then, that we do not have a separate word to describe discrimination that’s inherently wrongful?

KASPER LIPPERT-RASMUSSEN: This could be a disadvantage that, I think, in much theoretical literature can be easily remedied simply by adopting a specialized term—discrimination with a subscript or something like that—or simply stipulating that by discrimination, you will meet the following (e.g., differential treatment of members of different socially salient groups which is based on irrelevant reasons, or which is unfair), so I think for theoretical purposes it is less of an issue. However, I think that in much ordinary discussions about discrimination sometimes it’s a serious obstacle that people do not clarify that they actually have different notions of discrimination in mind.

How would the economists in the room define discrimination?

XAVIER RAMOS: I think that economists would subscribe to everything that Kasper just said. We can, however, try to add further comments on some of the issues. Certainly, what one can think of moralized or non-moralized concepts from our point of view is that concepts that are moralized allow one to guide policy action in a better manner. But it is often difficult to understand why some forms of discrimination are socially accepted and some are not, which is what Kasper was referring to. And discrimination is a great example.

I can think of two more relevant cases that may serve us to think about these issues. On the one hand, for instance, the price that men and women pay for car insurance is different ultimately because car accident rates are different. We (or our society) agree that, since risks are different, insurance companies are entitled to set different prices in order to make their business viable. On the other hand, social security contributions of men and women are usually the same across gender even though we know that the life expectancy of men and women differ. So, in this case, we do not allow discrimination by gender, whereas in the first case we do. I don’t know whether this has anything to do with people finding this treatment just or unjust or something else. But I also don’t know whether, in accepting this differential treatment, the fact that people conceive it as moralized or non-moralized or just or unjust is relevant.

On top of that, if we adopt a moralized concept, we tend to align moralized concepts with wrongful treatments or wrongful consequences, and this is not necessarily the case. Think, for instance, of affirmative action. Affirmative action or positive discrimination amounts to differential
treatment that aims to correct previous differential treatments and it is meant for a good cause. It is thus not supposed to be wrongful. So, it is not all together clear that the concepts of just or unjust serve to explain why people (do not) accept differential treatment. Moreover, it is definitely not clear that moralized equals wrongful.

*Let’s go to a deeper question about discrimination. According to some traditional accounts of wrongful discrimination, discrimination involves an objectionable mental state, which is rooted in prejudice or animosity. Many have argued that policies can be discriminatory without there being such mental states so long as they produce effects that are particularly bad for some group. Should we adopt such a concept of direct discrimination in our moral language? And wouldn’t this definition run the risk of being too inclusive?*

DIRK VAN DE GAER: It is an interesting question, but Xavier and I think that the concepts of objectionable mental states like prejudice or animosity are pretty useless. We don’t think they have much normative appeal.

The first reason is because we think about these kinds of objectionable mental states as a form of mechanism of approval or disapproval of certain behavior that serves to solve some coordination problems between people or to induce them to do something that is morally good. So, using these kinds of concepts (objectionable mental states) has no added value.

The second reason is that we also believe that there are forms of discrimination that do not involve any objectionable mental state. This is the case, for instance, of statistical discrimination for efficiency reasons, as opposed to self-confirming stereotypes. Statistical discrimination does not stem from prejudices or anything like that, and still, it is plausibly a form of discrimination, and that it is not good for society. In a nutshell, we see neither the value added nor the normative attractiveness of concepts like objectionable mental states.

*Kasper, do you agree with this statement?*

KASPER LIPPERT-RASMUSSEN: In my view, it all comes down, at least for certain purposes, to why one thinks discrimination is wrong. If, for instance, you have the view that discrimination is wrong because it is disrespectful and disrespect is tied to having a certain kind of mental state representing those people who may be discriminated against (as, say, having a lower moral status or being deficient in some other relevant way),
then, as far as I can see, it’s hard to get around talking about objectionable mental states—mental states that having those towards others involve disrespect. Of course, it’s true that it is a big issue when exactly a mental state is objectionable. But I think that’s simply an issue that is unavoidable, assuming that discrimination is wrong because it’s disrespectful.

However, I certainly agree with Dirk about there being forms of discrimination which are not disrespectful and, insofar as they are wrong, they are not wrong because of the objectionable mental states. So, for instance, if you think that certain kinds of sex discrimination are wrong because they bring about unequal opportunities for men and women, then that’s a fact about those forms of discrimination which does not depend on mental states.

One final point that connects to the previous issue is that if you define direct discrimination appealing to the notion of objectionable mental states, then that doesn’t make your concept of direct discrimination moralized. This is so because what defines a moralized notion of direct discrimination is that one can infer from the fact that something is direct discrimination that it’s morally unjustified. But the mere fact that one acts on the basis of an objectionable mental state doesn’t entail that the act in itself is morally unjustified, so it will still be a non-moralized conception of discrimination. I don’t think that goes against anything Dirk said.

*This is a nice bridge to the next question, which is whether or not objectionable mental states are involved in the definition of discrimination. Could it be the case that any form of discrimination is wrong because of its inequality producing effects? So, for example, because it increases group inequality or because it increases inequality of opportunity?*

KASPER LIPPERT-RASMUSSEN: My view is that there’s not a very tight connection between discrimination and equality. So, you can have forms of discrimination which most people would say are morally objectionable but nevertheless don’t result in any inequality such as, for instance, racial discrimination against the privileged racial minority. Most people would find that objectionable, but it would be racial discrimination that actually decreases inequality. Similarly, it is possible that there are forms of inequality which are not connected to what people would normally consider discrimination—say, inequalities between people who are naturally talented and people who are not, which would not normally be considered a
case of discrimination. But people might think that such forms of inequality are nevertheless objectionable inequality. Hence, I’d say that there is not a very tight connection between concerns for anti-discrimination and concerns for anti-inequality, even though, of course, one of the main objections to certain forms of discrimination is that they increase inequality.

XAVIER RAMOS: I agree with Kasper. The example of the ethnic minority illustrates well how these two concepts can be different and separate. Moreover, one exception to the thought that discrimination is wrong because it prevents at least some people from reaching their full potential (by, for instance, worsening the outcomes of the worst-off) is the one that Kasper was pointing out before.

Group inequalities, however, are different because they arise when people with exactly the same characteristics face the same hurdles to achieve their potential. So, group inequalities are an indicator that, because of certain characteristics, people are treated differently. Segall\(^2\) claims that discrimination is wrong only because it undermines equality of opportunity, which is a type of inequality. The reason is that it’s certainly unfair, which is what lies behind this relationship with inequality of opportunity.

But there are two further points. The first one is that some kinds of discrimination, for instance affirmative action, can actually equalize opportunities. These are different examples from the one that Kasper was referring to but go in the same direction. Second, discrimination attributes or grounds—characteristics on which discrimination is based, such as gender, race, ethnicity and so on—tend to be circumstances, i.e., characteristics that are beyond the control of individuals. However, if these discrimination grounds are responsibility characteristics (think for instance of obesity, which is the best example we can think of), they actually increase inequality of outcomes (for instance, when speaking of health) but not inequality of opportunity. In sum, when based on a characteristic for which individuals are deemed responsible, discriminatory treatment may increase inequality of outcomes but not inequality of opportunity. This example is in conflict with Segall’s argument.

---

\(^2\) Xavier Ramos refers to Segall (2012).
II. MEASURING DISCRIMINATION AND INEQUALITY

Let’s discuss the current state of the literature on measuring discrimination and inequality. How do we start thinking about this project of measuring inequality and discrimination? How do we operationalize these concepts, and what kind of tools are being used in your fields?

DIRK VAN DE GAER: In economics we have a very well-established tradition in measuring inequality and measuring poverty in which we formulate very precise properties that we want these measures to have. And then, based on those properties, we derive conditions under which we can order, let’s say, income distributions. So, we start from axiomatic properties and we combine them into a precise measure. This means that we try to define the concept as precisely as we possibly can in order to make clear statements about when one, let’s say, vector of income is more unequal than another one.

The economics literature on discrimination measurement relies on observational or correspondence studies to infer whether discrimination exists or not. This entails that one can only measure discrimination in situations like, for instance, the differences in the shares of individuals that are invited for a job interview or for visiting a flat or a house, between groups of individuals that differ in one characteristic. As a result, in terms of discrimination, we use empirical methods that do not really have a welfare basis: usually we look only at one particular relevant outcome (which is, in case of the example above for instance, whether a person gets a flat or not, whether one gets the job interview or not).

This is very different from the theory of inequality measurement, including the theory of inequality of opportunity, where very often we’re not only concerned with one particular outcome but we’re more concerned with something that is an overall measure of well-being, which may be income, but it could also be something else. In this sense, the way in which we measure discrimination nowadays is much less encompassing than the way in which we measure inequality.

Kasper, would you like to expand on this answer?

KASPER LIPPERT-RASMUSSEN: I’m beginning to move outside my field of expertise, so I just want to add that I think it might be interesting to look to other issues comparable to discrimination (say issues about altruistic versus non-altruistic behavior) and ask whether there’s any kind of overall measure for the degree to which people act altruistically or non-
 altruistically in society. I would suspect that there’s very little precise measurement of that sort, and I think that if we want some kind of measure of how much discrimination exists in a certain society, many of the same obstacles that economists have come up against would come up again in an instance like that. So, one can measure the degree to which discrimination takes place in a particular sphere, against a particular group, but the sort of overall measurement about how much discrimination goes on in the society, I suspect that that’s a hugely complicated thing to measure.

Do philosophers have an important role in formalizing a concept of discrimination that is easier to measure by practicing scientists?
KASPER LIPPERT-RASMUSSEN: I don’t have a good overall answer to that question. I’m involved in two experimental studies at the moment where we try to measure the degree to which people’s inclination to classify something as discrimination depends on the existence of a comparator who’s been treated better and, secondly, whether or not people object to discrimination based on whether it involves disrespect or harm. And in both studies it’s very hard to ask the sort of questions that philosophers would like to ask in order to test people’s intuitions. I suspect that similar issues would come up in a lot of attempts to measure the degree to which discrimination takes place.

I think the main contribution of philosophers in this respect would probably lie in trying to provide some more analysis of the concept of discrimination and not so much in the effort to operationalize the concept of discrimination, which I took was what described in relation to having these very precise concepts of discrimination that economists measure.

When it comes to measuring discrimination and inequality, there might be some trade-offs. One trade-off we thought economists might deal with is the following: on the hand, they want to gain society-wide information on the extent of discrimination and inequality, even if the quality of research might be really limited but, on the other hand, they might want to conduct smaller scale studies that allow for precise insights into the structure and nature of discrimination of inequality, though this may be less useful to measure the extent, or societal impacts, of discrimination and inequality. So, do you think that there is such a trade-off? And what are the benefits and limits of either option?
XAVIER RAMOS: We realized that often, at least in inequality measurement, the trade-off is not between the width of the group we want to study (that is, for instance, whether we want to study a population or group-inequality in that group or in the entire population) and the quality or the validity of the research conclusions. This is different from the case of measuring discrimination, where there are a lot of case studies and we may think that often these are either informative or externally valid.3 The reason why, when we analyze income inequality, this is not so much the trade-off that we face is because studies suffer from different types of problems. Let me name just three of them.

First of all, we lack information about both tails of the income distribution. With the data we have at hand (basically survey and administrative data) it is difficult to collect data on very poor people (those, for instance, that are living under a bridge or in collective houses like prisons) and also on the top part of the distribution, although there are large efforts being expended by quite famous economists like Thomas Piketty or Tony Atkinson (who passed away shortly ago). So, this problem applies to these ends of the distribution and not to particular subgroups of the population defined by different socio-economic characteristics.

The second problem is that the degree to which using some income sources can be applied to the analysis of other classes of income is limited. One famous example is that of self-employed earnings. We know that self-reported earnings are not truthful, so we are capturing this type of earnings incorrectly. And we encounter self-employed individuals in each population subgroup.

Thirdly, there is also lack of information about important behaviors when we think of income inequality such as tax evasion or tax avoidance, which prevents us from measuring inequality properly. Once again, this is very much related to income but perhaps not so much to all other characteristics. In addition, we know that different sources of data that we can use in order to measure income inequality (I have named two of them—survey data or administrative records—and currently there is huge work on tax records and on social security records) are captured at the individual level or micro level and they do not match with national account data. It is on these latter data that many policies are made. There is, again, a huge effort by the same group of scholars at the Paris School of

3 The external validity of a study indicates the extent to which it is possible to generalize the results from one case study to another relevant subgroup or other cases of discrimination.
Economics that I mentioned earlier to make progress to match these tools, but these data problems apply to all population’s subgroups.

These problems apply differently to different parts of the income distribution, but they may apply to different population subgroups. So, there is often no benefit in trying to focus on a population subgroup and analyze the subgroup with the expectation that we are going to have better data and therefore better measurement. The reason, as I have explained, has to do with the income data problems that we face. The take-home message is that the strategy of focusing on a group may be informative and useful for discrimination analysis but not so much for inequality analysis.

Let’s return to discrimination and inequality of opportunity. The measurements of discrimination and equality of opportunity are currently seen as distinct research projects, even though we think there is considerable conceptual overlap between the concepts. So, should we move towards greater integration between these measurement projects or, if not, why not?

DIRK VAN DE GAER: Concepts need to be defined very precisely. Kasper already said that in the context of discrimination there is a disagreement about what people mean by discrimination and the same is true with inequality of opportunity—there are very many ways in which these concepts can be defined. And I think that the main focus should be on getting very precise definitions of these concepts and all of their different variants, and then investigate their implications for ordering different possible societies (a ‘society’ is here understood as a complete description of society).

By taking each one of these different conceptualizations separately, one will get one particular conceptualization or one particular variant of a concept and, from that, one can deduce certain conditions under which one can determine whether a society is better than another one or not. It will not always be possible to order societies since the ordering will likely be incomplete. This latter problem, however, can be addressed by increasing the precision of the conceptualizations (or variants), which makes the ordering more complete. I think that this should be done for every possible conceptualization and every variant separately and then confront the conditions for ordering societies. This is the only way in which we can then establish whether the different variants or different conceptualizations will agree on the way they rank different societies.
This looks like a much more fruitful way than starting from the concepts and trying to narrow down what is the common theme of the concepts, which I don’t think to be a very fruitful exercise. I think it’s better to try to work with as precise conceptualizations as possible, apply them consistently, and then then look at the conditions for ordering societies that one can derive from these very precise concepts, even if they do not determine a complete ordering of all societies.

III. THE RELEVANCE OF MEASURING DISCRIMINATION AND INEQUALITY

*We have discussed many conceptual questions. Let us now bridge the gap between the conceptual questions and the policymaking applications. Why, if at all, is it important to measure discrimination and inequality? It seems a very hard and controversial process. Why is it important to have a quantitative understanding of this phenomenon, rather than rely on a purely qualitative one?*

XAVIER RAMOS: Both things are relevant, and the latter question is more relevant for policymakers. Detecting that there is discrimination or inequality is informative and helps, but what can policymaker do if they only have such coarse-grained pieces of information? Should they put lots of resources into addressing inequality or discrimination, or very little? What is the part of the government budget that they should want to devote to that?

Quite likely, this depends on many issues such as social relevance, or even social alarm, but also on whether one thinks that this is a huge or a small problem. This, in turn, will depend on the extent to which one believes that there is inequality or discrimination. The mere presence of discrimination or inequality can help one take some decisions, but certainly not design policies that specify the amount of resources or effort devoted to address the problem, especially knowing that all the resources and efforts that we have are scarce and devoting them to address inequality or discrimination implies subtracting them from something else (whenever resources are employed for something, there is always the opportunity cost of not doing something else). As such, measuring the extent of the problem is extremely relevant and we cannot do without that. I think we are going to be more helpful to policymakers if we can tell them that discrimination is large or small. And how large and how small this is.

Not all methods are able to say that. For instance, the correspondence analysis that Dirk was mentioning before, which economists use in order
to detect discrimination, often cannot say much about the extent to which there is discrimination. By sending letters with invented CVs and names or other characteristics, it is possible to detect that there is discrimination because some get more job interviews than others, as Dirk was saying. And yet, it is not possible to gain a deeper knowledge of the issue, since it is not possible to fake the interview in a relevant and credible way. Measuring the extent of discrimination is therefore crucial.

KASPER LIPPERT-RASMUSSEN: I guess the reason why the question is slightly provocative is that it somehow implies that it’s an either-or question, but, of course, I suspect that the best thing would be to rely both on quantitative measures and qualitative measures of discrimination. In relation to this, a supplementary reason why it might be good to have quantitative measures is that often there can be a gap between, so to speak, experienced discrimination and actual discrimination, so if the qualitative judgments are judgments somehow reporting or being constructed on the basis of when the people themselves see themselves as being discriminated against or discriminating, then those judgments are likely to be quite misleading in a way which, say, making precise quantitative measurements of discrimination can actually show to be misleading.

It seems that the connection between scholarly work on both the concepts of discrimination and inequality and its measurement and policymaking can be complex, or at least not straightforward. Let us thus ask two questions. On the one hand, if there were no connection between policymaking and scholarly work, would it still be relevant for scholars to try to pursue this project of measuring discrimination? On the other hand, what would be the role of scholars in cases in which the connection between their work and policymaking were noisy (their research outputs were used superficially or misunderstood by policymakers)?

DIRK VAN DE GAER: I think that the problem is that the task of detecting discrimination and inequality is not only to inform policymaking, it also allows us as economists, for instance, to study how markets work and how inequality arises in the economy. It’s important to understand human behavior, and this has a much wider relevance than just trying to inform policymaking, so its task is wider than just informing policymakers. What happens when policymakers misunderstand core concepts?

Well, I think that our gut feeling is that society will be better off if policymaking is evidence-based on scholarly work. And I believe that there’s a quite a bit of progress nowadays: scientific communicators or
properly trained technicians and advisors can help policymakers to gain a correct interpretation or conclusion. With the right kind of communication, much can be obtained. Moreover, sometimes policymakers might not even want to understand what we’re saying. In some cases, they do so because it’s a way to keep a free hand and to implement a policy that they like for ideological or other reasons, rather than implementing a good scientifically based policy. This notwithstanding, we should try to communicate as clearly as possible, not just to policymakers, but also to the general public.

KASPER LIPPERT-RASMUSSEN: I think that perhaps the relevant comparison is whether it’s better that policymakers act on the basis of various misunderstandings of research results, rather than act on the basis of no information about research results at all. In my view, it might well be better to act on some deficient understanding of research results than acting on no research results at all.

Before wrapping up the session, I’d like to ask whether there are questions from the audience.

AUDIENCE: One of the things Professor Van de Gaer was talking about was that we want characterization results to understand the phenomena properly. Specifically, characterization results of distinct conceptualizations of inequality. But a problem is that these conceptualizations are welfarist: they rely only on vectors of well-being or income or some sort of understanding of preferences, which allow us to get the ranking of distinct vectors of different states of affairs. However, discrimination is relational. It goes away from the question of ‘equality of what kind of thing’ to a kind of game-form understanding of ‘is someone else frustrating the opportunities that I confront?’. So, do we need to move to game-forms to understand discrimination, rather than focus on characterization results that we use in inequality measurement?

DIRK VAN DE GAER: Do you want to use game-forms for the purpose of understanding discrimination or measuring discrimination?

AUDIENCE: The purpose of it is to get some sort of conceptual grip on the phenomena, since we take discrimination to be some sort of opportunities sets that are not available for another person. Insofar as these

---

4 The question was asked by Akshath Jintendranath, a PhD student at the Vrije Universiteit Amsterdam and a former editor at EJPE.
opportunities sets are jointly produced, should we be using more game forms in economic theory to understand the phenomenon of discrimination, rather than welfare functions (which economists use for inequality)?

DIRK VAN DE GAER: I'm not too sure what we should do that in terms of measuring. However, this is very important for understanding how discrimination arises. We still have a lot of problems in incorporating these kinds of phenomena which arise, of course, in game-forms, which can be understood as externalities. It is very difficult to deal with that from an evaluation formal point of view. As such, everything that we do in terms of measurement theory is very individualistic. We do not have a clear answer to that, at least yet, I believe. So, all the rankings of different states of affairs in terms of discrimination or inequality are based on individual outcomes, and therefore are 'stipulated' from one individual perspective or another. How the outcomes ranked arise is a positive question. However, I believe that some of the discrimination models that have to do with stereotypes can be given a game-theoretic interpretation, but there's a distinction between the normative side and the positive side. And so, in terms of evaluation purposes, we're still very individualistic. But on the positive side, there's lots of things that we can do, and game theory could indeed be very relevant as an approach to get things going.

IV. THE FUTURE OF MEASURING DISCRIMINATION AND INEQUALITY

Let us wrap up the roundtable with a question that looks at the future of the field. How do you envisage the development of the field, both of measuring discrimination and the conceptual work on what inequality and discrimination are? In what ways could these projects improve? Is there room from increased collaboration between different fields? Is there something to be gained by interdisciplinarity?

XAVIER RAMOS: It is a difficult question. And so, precise answers are difficult as well. Let us say that we would like to see the field going much in the same direction that Dirk mentioned in one of his answers. We believe that the best way forward is to acknowledge and identify these different ways to conceptualize concepts such as discrimination and inequality as precisely as we can, recognize them as different concepts because the more precise the definitions we’re going to have the more social states we’re going to be able to order. And then, if there are contradictions because the concepts are incompatible, we’ll have to make a choice.
After all, this is often what happens when one tries to use different tools to measure inequality. In such cases, different tools correspond to different sets of axioms that the measure should respect, which then define the basic properties that characterize the measure. But it often happens that using different tools (i.e., different axiomatizations) leads to measures that have basic properties that clash. Because the underlying properties are clashing, the economist using the measure has to make a choice. Since the clashes between measures arise from disagreements about the characterization of the concepts used, the primary issue that needs to be addressed is conceptual in nature. Whether economists can measure discrimination or inequality from these properties or not is a different problem, and it’s more technical.

But the way forward is to try to identify as precisely as possible the different meanings of what is commonly known as the same concept. In this regard, the collaboration between different social scientists or philosophers sometimes is (and has proven to be) key and a good way to proceed forward. Equality of opportunity is a great example. It’s a field that started in philosophy, from which economists drew. With their tools or interpretations, they attempt to push it forward. If we are at the stage we are now, in which we are able to characterize different possible concepts of equality of opportunity with different measurement tools, this is because this collaboration went through. It often helps that there are scholars with one foot in one field and one in the other, like Marc Fleurbaey or John Roemer, who serve to bridge the two.

KASPER LIPPERT-RASMUSSEN: One area of conceptual work needed in the future is work that explores what makes differential treatment of certain groups into discrimination. Historically, sex discrimination, race discrimination, and religious discrimination were the, so to speak, paradigm cases. But it’s clear that the list of groups that are considered victims of discrimination has expanded considerably since the civil rights movement in the ‘60s in the US (think about discrimination against smokers or appearance discrimination). So, a better understanding of the concept of discrimination in terms of what the relevant properties are such that differential treatment on the basis of having those properties qualify that as discrimination.

Besides that, I suspect that there are some interesting issues in relation to poverty. Generally, people do not think of poor people as a group that is subjected to discrimination in various forms. Compare for instance the preference that upper-middle class people have for avoiding living in
areas where poor people live in part because of the socio-economic status of people who live there (in part for other reasons) with the preference many racial-majority people have for avoiding living in areas where a racial minority form the local majority in part because of the racial identity of most people who live there (in part for other reasons). I suspect acting on the former preference would generally not be considered engaging in poverty discrimination, whereas acting on the latter preference would generally be considered engaging in racial discrimination (albeit permissible racial discrimination perhaps).

I believe, however, that, under many understandings of what makes differential treatment discrimination, it’s hard to explain why these two cases should be regarded as different from the perspective of discrimination. If they shouldn’t, this can have quite radical implications. For instance, it would imply that we should adopt a view of standard social mechanisms of socio-economic segregation which is as critical as our views of racial segregation, since in the end both are forms of morally objectionable discrimination.

In terms of measurement, I think once one has a better understanding of the list of properties which are such that differential treatment of people who have these properties can cover discrimination, it might also be interesting to see if some kinds of discrimination that we classify in one way should at least to some degree be classified in a different way. Think about the relationship between race discrimination and appearance discrimination. Presumably there might be cases which are now classified as race discrimination which might be better conceived as cases of appearance discrimination. The kind of cases I have in mind here are cases where people treat people differently on the basis of their facial appearances differing in racially stereotypical ways irrespective of whether these people are considered to be members of the same race, whatever that is. So, I think that there are some interesting things here to explore.

REFERENCES


JON MURPHY
Western Carolina University

By 1950 a half-century of political and social events seemed to suggest that the liberal mindset of limited government and promotion of individual rights that had generally dominated Europe and America during the 19th century was dying. World War I and the Great Depression seemed to undermine the cosmopolitan liberal idea of doux commerce. The intelligentsia of the day were teaching and promoting socialism, eugenics, and other technocratic ideas. Two totalitarian threats, German fascism and Japanese imperialism, had been defeated on the battlefields of Europe and Asia only to be replaced by the specter of communism. In the United States, ‘illiberal liberalism’ was taking hold as Senator Joseph McCarthy’s used the threat of communist infiltration of the United States government to justify prosecuting political opponents.

How would liberals be able to defend liberalism without abandoning the ideals of liberalism? That is the key question Joshua Cherniss seeks to explore in his book: “How to combat anti-liberal movements, which are not constrained in the way that liberal movements and regimes are, without sacrificing political efficacy or betraying basic liberal principles in the name of defending them?” (5). To answer this question, Cherniss looks at four major liberal thinkers of the 20th century: Albert Camus (chapter 3), Raymond Aron (chapter 4), Reinhold Neibuhr (chapter 5), and Isaiah Berlin (chapter 6). In particular, he seeks to explore their ethos, a “‘stance’ or ‘bearing’ formed by patterns of disposition, perception, commitment, and response, which shapes how individuals or groups go about acting politically” (6) to synthesize it into a liberal ethos. Cherniss wants to re-align liberalism around a tempered ethical center, as opposed to a formal institutional structure or set of general principles: How can one be a good liberal as opposed to merely speaking like one (6–7)? This leads to one of his major claims: “There is something not only paradoxical, but (potentially) self-defeating, and even pathological, about seeking to live a liberal ‘creed’ through an illiberal ethos” (7). To defend liberalism means not only
supporting liberal institutions, but to treat people in a liberal way. Cherniss argues that promoting a liberal mindset is important to defending liberalism, both in the post-war period he analyzes and in light of contemporary challenges to liberalism.

Cherniss is not attacking specific illiberal systems, but developing a ‘tempered’ liberal ethos and arguing against an illiberal ethos. This represents his major contribution to the literature. One will not find many references to fascism, socialism, imperialism, or other ideologies within this book except insofar as they represent the various governments under which people were living in the 20th century. Rather, he “criticizes a set of impulses often exhibited in anti-liberalism: intolerance, self-righteousness, craving for simplicity and certainty, deafness to dialogue, righteous ruthlessness” (13). These impulses serve to tempt us into an “anti-anti-liberalism resembling that which it opposes in dogmatism, self-righteousness, and intolerance” (13). Camus, Aron, Neibuhr, and Berlin were tempted into abandoning the liberal ethos in order to save it, but Cherniss discusses how they resisted. This is a book written for liberals about liberalism.

Much of modern research explore liberal institutions rather than ideals of liberalism. For example, liberal political economists such as James Buchanan discuss at length necessary institutional arrangements (‘constitutional political economy’) to protect and promote a liberal society. Debates within liberalism tend to occur over the proper role for government in social and economic arrangements. Political thought tries to define concepts such as liberty, equity, justice, and the legitimacy of political institutions and policies. However, without denying the importance of these conversations, Cherniss develops a narrative on the mindset of liberalism. His major contribution to the literature is developing this ethos, or stance, that characterizes liberal thought, rather than try to tease out universal ideals or desirable political institutions. In other words, rather than focusing on constitutions or the concept of the rule of law as the foundation of a liberal society, Cherniss attempts to reorient the conversation around a liberal mindset to preserve these institutions as liberal.

By focusing on the liberal ethos, Cherniss provides a new lens through which to view political conflict and debate in the 20th century. The Cold War was just one aspect. The liberal thinkers covered in this book were battling for the very soul of liberalism as well. Defeating the communists would be an empty victory if we liberals merely adopt their anti-liberal methods. Consequently, we can better understand the development and
challenges faced by liberal thinkers in the post-World War II years. Additionally, I think we have a framework to understand the strong liberal desire following the fall of the Soviet Union to promote peace through trade. The economic policies of much of the world in the 1990s and 2000s suggest that the liberal ethos won the day, not merely the facade of liberal institutions.

The book can be divided into three parts. Part 1, which comprises chapters 1 through 2 motivate Cherniss’ argument for a liberal ethos. Part 2 consists of chapters 3 through 6. These chapters act as mini biographies of the liberal thinkers: their struggles in a tumultuous world, their interactions with other thinkers and each other, and how political realities shaped their philosophies. Part 3, the conclusion chapter, attempts to synthesize the lessons of the 20th century, the tensions within liberalism that Camus and the others uncover, and apply the lessons for the 21st century.

The focus on an ethos is necessary for the story Cherniss wishes to tell. Quoting from Bertrand Russell, he tells us “the ‘essence’ of liberalism ‘lies not in what opinions are held, but in how they are held’” (35, emphasis in original). Inversely, anti-liberalism “was defined not just by an alternative model of political institutions, but also by an ethos of discipline, dedication, resoluteness, and ruthlessness, which rejected traditional (liberal) moral standards as signs of weakness” (35). Ethos helps us understand the fervor of anti-liberalism. Fascism, communism, and Naziism did not represent merely different allocations of political power for their supporters. They were struggles of good versus evil, life versus death, progress versus stagnation. Liberalism was not merely differing opinions on government. Liberalism was a moral failing in the eyes of the anti-liberals. Consequently, just about every tool was open to anti-liberals to crush it (36–37). During the 20th century, liberals were besieged by anti-liberalism and had to reorient liberalism to address its weaknesses, develop its temper, and exemplify the liberal spirit (38). Thus developed the ‘tempered liberalism’ of the 20th century.

In this book, ‘tempered liberalism’ represents two things: first, the historical position of the aforementioned liberal scholars; second, a normative position that Cherniss advocates in the conclusion (198). It is a character and method of analysis that is “opposed to systematization, insistently independent of doctrine, and reliant on discriminating judgment” (198). Skepticism is the order of the day, both external and internal. Tempered liberalism matches closely with Adam Smith’s warning against
the Man of System, who becomes enraptured of the beauty of his plan at the expense of his humanity (Smith [1759] 1982, 233–234). Like Smith, Cherniss’ tempered liberalism is aimed both at opponents of liberalism and at supporters of liberalism. Just as we are skeptical of the grand schemes of the anti-liberals, we must be skeptical of our own plans.

Self-restraint plays a key role in the mind of the tempered liberal:

They [the liberal scholars discussed in the book] called for fortitude in tolerating others—which means mastering tendencies toward disapproval, irritation, and the desire to step in and take control when others are making a mess of things. Liberal fortitude is often a matter of exercising forbearance—a disposition, posture, and practice that refrains from exercising power, or exploiting one’s advantage, over others. (199)

Liberal forbearance is tempered by humanity and Smithian sympathy: “The perception of others’ humanity, and the sense that those others are fellow creatures who call forth sympathy and a basic degree of respect” (199).

This theme of humanity and sympathy is ever present throughout the book and the thinking of the scholars. In that sense, this book follows thinkers (including Cherniss) very much in the tradition of Adam Smith. In the discussions of each thinker, one can spot the liberal influences of Adam Smith. For example, on Camus’ discussion of justice, Cherniss writes: “The pursuit of justice should be balanced by other moral considerations, and by a love of life that goes beyond morality, nourishes it, and prevents it from becoming doctrinaire” (76). Justice is a method of thinking, not so much a state of being. That the virtues temper each other is a major theme throughout Adam Smith’s *Theory of Moral Sentiments*.

In the conclusion, we are treated to Cherniss’ own thoughts on tempered liberalism. He argues that we are experiencing many of the same issues as the post-World War II world:

We have of late seen of increasingly naked, cynical ruthlessness, which makes fewer concessions to liberal norms—and the embrace of such ruthlessness among leaders (and rank-and-file) of mainstream parties in supposedly stable liberal democracies […]. The appeal of strongmen and bullies is still potent, as is the lure of all-encompassing, supposedly infallible ideologies, and the pressure to take sides, and prosecute the struggle furiously. (219)
Both the Left and the Right (one assumes he is referring to American politics) are provoking each other into “righteous, energizing fury” against liberal institutions (219). Thus, his analysis is timely. We can look to the liberals of old for hope.

There are no specifics on how to apply the lessons of tempered liberalism to contemporary social and political problems. But, as I say above, that is a feature rather than a bug of *Liberalism in Dark Times*. This is a book about the liberal mindset, as opposed to anti-Trumpian conservatism or anti-Progressivism. Rather than lecture on specifics, Cherniss exults us to sustain “the will to fight for liberalism” and maintain “awareness of the reasons it is worth fighting for” (220). When we focus on fighting this or that policy, personality, or party, we may lose our ethos of liberalism and descend into anti-liberalism ourselves. One must not fight evil by allying with the Devil. When one deals with the Devil, the Devil always wins.

In the conclusion, Cherniss presents us liberal academics and educators with a challenge. How do we instill this liberal ethos in ourselves and our students? How do we internalize the lessons of Aron, Berlin, Cherniss, and the others so that we can reorient our societies around tempered liberalism? How do we forge a tempered liberal ethos within ourselves? Cherniss leaves that question as an exercise to the reader. Indeed, the exercise must be left to the reader as the preceding chapters show: all the liberal thinkers faced different political and social problems. The uniting bond was the tempered liberal ethos, but the specifics of each thinker’s response differed. Liberalism is about this ethos, rather than specific constitutions or institutions.

There is one criticism of the book. Cherniss sets out to defend the liberal mindset from illiberal threats, both within (for example, ‘How can liberalism win when our enemies use illiberal tactics?’) and without (for example, ‘Liberalism has failed. We need a new system’). His focus is on threats from within. He is generally successful in his goal. I was convinced after reading of the virtues of a liberal mindset guiding our activities, as opposed to simply advocating for liberal institutions. However, I fear those already not sympathetic to a ‘tempered liberal’ mindset will reject his arguments. In that way, Cherniss faces the same problems as Camus, Aron, and the others as they sought to overcome the illiberalism of the 20th century. I suspect the book will not have the direct impact of shifting popular political sentiments back toward liberalism. This book may not win over the hearts of illiberal readers, but it does build a strong ethos...
from which readers of Liberalism in Dark Times can engage illiberalism and strengthen our resolve.

Those of us not trained in political philosophy will find some of the discussion difficult to follow. However, Liberalism in Dark Times is well worth the investment of time. Liberalism is facing another crisis, just as we did nearly a century ago. After the fall of the Soviet Union, liberalism rested too much on its laurels, and, like the liberalism of the early 20th century, perhaps became a little decadent. But this book provides a beacon of hope: liberalism has long been tested by illiberalism and our predecessors faced significant illiberal threats from around the world. Despite the darkness, they kept to liberalism and helped it prevail in the late 20th century. We in 2022 are not unique in our struggles and the lessons of the 20th century tempered liberals provide insights into how to weather this storm: know when to be headstrong and when to be soft-hearted, as this will help keep one tempered without becoming manipulated. Respect the dignity of all, from the oppressed to the oppressor, as this will keep one from slipping into vengeance. Know why it is we support liberalism: the basic dignity of all. This is how liberalism survived the illiberal onslaught of fascism and socialism in the 20th century. And this is how we liberals must respond to the issues of the 21st century. To paraphrase David R. Henderson: liberalism is a hardy weed, not a delicate flower.

REFERENCES

Jon Murphy is an Instructor of Economics at Western Carolina University. Jon’s research focuses on both the production and consumption of expert opinion in government, business, and personal operations. He has also written on the history of economic thought, specifically in relation to Adam Smith. He holds a Ph.D. in Economics from George Mason University. Contact e-mail: <jmurphy8289@gmail.com>

VAIOS KOLIOFOTIS
Eindhoven University of Technology

In the last few centuries, large parts of the world have become wealthier. According to standard economic theory, technological progress is one of the main contributors toward economic prosperity. Other important factors are institutional rules that protect property rights and promote innovation and competition among producers of goods and services. While the literature on institutions within economic science has grown immensely in recent years, theoretical and empirical research on the relationship between power and institutions is sparse. Building on decades of research on economics and public policy, Randall Holcombe’s recent book Coordination, Cooperation, and Control looks into this neglected topic and focuses on power as a key concept in explaining the origin and evolution of institutions over human history.

According to Holcombe, one of the main reasons for the increase in wealth in the last centuries is the separation of economic and political power. The main claim of the book is that when the same people hold both economic and political power, the result is stagnation. Conversely, when economic and political power is separated, the result is economic progress. Therefore, keeping those with economic power from obtaining political power, and vice versa, is necessary to maintain the economic prosperity the world has seen in the last few hundred years.

In more detail, the 13 chapters of the book can be roughly separated into three parts. The first part includes four chapters and is devoted to examining in detail the properties of political and economic power. The concept of power is first presented in chapter 1, while the characteristics of political power and economic power are analyzed in chapters 2 and 3. Finally, chapter 4 examines how the separation of political and economic power occurs.

As Holcombe explains in the first part of his book, economic progress is only possible because of the coordination of the economic activities of individuals from all around the world. This coordination can oc-
cur in two different ways: cooperation and control. Markets work on the basis of cooperation between buyers and sellers and the voluntary exchange of goods and services. In contrast, governments operate by control, threatening to impose costs on those that do not pay taxes or abide by regulations of government institutions.

The concepts of economic and political power examined in this part have an institutional foundation. According to Holcombe, economic power is institutionalized control over resources exercised through voluntary cooperation with others. In contrast, political power is institutionalized control over the actions of people through threats of coercion (force and sanctions) against those who do not comply. However, cooperation in market exchanges requires an institutional framework that clearly defines and enforces property rights. This is created and protected by those who have political power.

While in the first part of the book Holcombe does not offer indisputable definitions of political and economic power, his way of presenting those concepts provides useful insights for understanding how they influence economic prosperity. When economic and political power is separated, those who have economic power have an incentive to engage in profitable activities and look for ways to create value for others. This results in more efficient production processes and new and improved goods and services. Therefore, when economic power is exercised independently of political power, the result is economic progress. In contrast, when people who hold economic power also acquire political power, the result is economic stagnation. Instead of promoting market competition and innovation, they use their political power to provide themselves with economic benefits through tax cuts, subsidies, tariffs, and market entry barriers.

The first four chapters provide the conceptual foundation for analyzing political and economic power. The second part consists of four chapters that build on this foundation to examine the evolution of economic and political power from a historical perspective. The chapters within this part address economic and political power in pre-agricultural societies (chapter 5) and in agrarian and feudal societies (chapter 6), the characteristics of institutions that support commerce and industry (chapter 7), and the power relationships developed in commercial and industrial societies (chapter 8). One of the main claims here is that while economic power has almost always been dominated by political power, the remarkable increase in economic prosperity in the last few hundred
years has been the result of the development of economic and political institutions of capitalism that led to the separation of economic and political power.

In more detail, Chapter 5 draws on a plethora of evidence to describe societies before the Neolithic (first agricultural) revolution. A key idea in this chapter is that in pre-agricultural societies the most important factor of production is labor, and that different types of power are not clearly differentiated. All relationships were based on labor activities and the personal knowledge and reputation of group members. Different types of power are combined because all of them are exercised over the activities of individuals. Although it is often difficult for individuals to leave pre-agricultural societies, groups can exit and take their labor with them. This ability provides some check on the abuse of power.

In chapter 6, Holcombe examines societies after the agricultural revolution where land displaces labor as the most important factor of production. In these societies, economic power is based on the possession of land that brings with it political power over those who work on that land. Therefore, in agricultural societies, political and economic power tends to be in the hands of the same people. There are also fewer checks on the abuse of power because those without land cannot easily move away from agrarian societies. Holcombe suggests that, because of the combination of economic and political power, societies were stagnant and economic progress was very slow.

In chapters 7 and 8, Holcombe examines modern industrial and commercial societies where capital and entrepreneurship are the most important factors of production. These societies are more productive than pre-agricultural and agricultural societies, creating economic surpluses. Capital and entrepreneurship are mobile and can therefore move and escape the abuse of political power. Moreover, capital depreciates and requires constant investment for its maintenance in a particular area. Therefore, there are incentives for those with political power to protect and tax those with economic power. Holcombe concludes that commerce and industry can facilitate the development of institutions that allow economic power to flourish independently from political power.

After a detailed historical analysis in the second part of the book, the third part examines topics familiar to followers of the Public Choice tradition. Holcombe’s analysis of power provides some new ways to examine these theories and issues faced by contemporary societies. Chap-
ter 9 analyzes politics as a vocation, chapter 10 re-examines social contract theories, chapter 11 focuses on the role of ideology on the separation of political and economic power, and chapter 12 examines issues related to the recombination of political and economic power. Finally, chapter 13 applies ideas developed throughout the book to examine economic progress in countries around the world.

In chapter 9, Holcombe examines the differences between plunder and taxation, the role of legal institutions, and the main characteristics of political competition. One of the main ideas of this chapter is that governments are based on the exchange of protection for tribute. To practice politics as a vocation, it is important that there is a surplus of economic resources that can support those that have political power. In commercial and industrial societies, instead of engaging in plunder that provides only short-run benefits, those with political power have an incentive to collect regular tribute over a long time. To avoid plunder, producers of economic wealth have an incentive to accept the legitimacy of those that hold political power and provide them with economic resources over a long time-horizon.

Chapter 10 of the book can be divided into two parts. The first part discusses social contract theories from political philosophy and relates them to the discussion of power in the previous chapters. According to Holcombe, a plausible way to view the social contract is as a set of generally accepted constitutional rules that restricts the government and determines the citizens’ rights and obligations. In the second part of chapter 10, Holcombe discusses a way of measuring the separation of economic from political power. The Fraser Institute’s Economic Freedom of the World (EFW) index is designed to provide a measure of the level of economic freedom around the world—the degree to which people are free to engage in voluntary exchange with others and achieve mutual benefits. In this chapter, the EFW index is also used to measure the degree to which economic and political power are separated.

In chapter 11, Holcombe examines public opinion and the production of ideas as important factors for the separation of political from economic power. According to Holcombe, democracy can be viewed in two main ways. One view of democracy is as a mechanism for determining who holds political power and enables citizens to peacefully replace those who exercise the power of the state, if they desire to do so. Another view, expressed by Jean Jacques Rousseau, depicts democracy as a revealing the public interest. According to Holcombe, this second view
of democracy depicts the actions of a democratic government as carrying out the will of its citizens and justifies government redistributive policies that impose costs on some for the benefit of others.

In chapter 12, Holcombe examines the tendencies of those with political power to try to use it to gain economic power, and vice versa. Public choice theory provides the details of the mechanisms through which the recombination of economic and political can occur. One key point here is that regulation does not only constrain but solidifies the position of those with substantial economic power by protecting them against competition. Over time, profit becomes increasingly dependent on political connections rather than the production of goods and services for consumers. Another important idea borrowed from public choice theory is that those with economic power can effectively organize to gain benefits for themselves. They can gain benefits by entering a low transaction costs group that negotiates with those that have political power. These groups tend to be small because the economic benefits to each individual group member are larger and the transaction costs to organize into such a group are lower.

The last chapter of the book (chapter 13) summarizes the main arguments developed in the previous chapters and applies the new insights to contemporary societies. According to Holcombe, political and economic power can remain separated if political power rests with the positions and not the personal characteristics of those that hold those positions. Moreover, institutions that create checks and balances can constrain the abuse of power while those socialist ideas that advocate the recombination of economic and political power have to be treated as a threat to economic prosperity. In this chapter, Holcombe also applies the Fraser Institute’s EFW index, a measure of the degree to which people are able to exercise economic power independent of political power. It is argued that many countries, such as Singapore and Georgia, conform to the main claim of the book that separating economic power from political power results in economic progress. In contrast, the case of Venezuela demonstrates that when those with political power confiscate economic power the result is economic stagnation.

There are issues with the specifics of some of the arguments developed in the book that I will examine below. However, the analysis that Holcombe promotes is relevant and novel. An examination of the institutional sources of power is not typically undertaken by economists. Moreover, Holcombe’s arguments on the separation of political and eco-
nomic power can be fruitfully applied to examine many developed societies. For example, Holcombe’s analysis can be used to explore the interaction between those with political and economic power in countries like Greece. One can find here a long history of co-dependence between political dynasties and families with economic power that goes all the way back to the days of the Ottoman Empire. Although this phenomenon is not unique to Greece, or other Balkan countries, the case of Greece provides a clear example where political and economic power are combined and can be used for the application of Holcombe’s analysis.

While acknowledging Holcombe’s important contribution, this book also has its shortcomings. The first weakness involves some of the empirical research presented in the book. The discussion of past societies in the second part is well-informed by the latest historical and anthropological evidence, and Holcombe puts forward some thought-provoking conjectures about power which appear to be consistent with the historical development of economic and political institutions. However, in the third part of the book, it is argued that the EFW index can be also viewed as a possible measure of the degree of independence between economic and political power. Leaving aside the definition of economic freedom on which this index is based upon, the results—when the index is applied in the last chapter to examine power relations in Asian countries—are far from satisfying. As Holcombe admits, countries such as China demonstrate that there is evidence that economic progress can occur even if economic and political power is combined. Among many factors, what explains the economic prosperity of countries that are ranked low according to the index is the low cost of labor and the import and further development of technology from other countries. Holcombe suggests that these advantages are temporary, and the result will end up being stagnation. However, at best this is a conjecture, and Holcombe does not provide any evidence for establishing its veracity. It is, I think, more plausible to suggest that research on economic growth and the variability of political systems demonstrates that there is no evidence for linearity in the origins and development of economic prosperity. Different societies appear to be on their own path to economic progress and the main claim of the book—that the combination of economic and political power leads to economic stagnation—does not appear to be correct for all developing countries.

Another problematic aspect of the book is that, despite Holcombe’s work of clarification, power remains a contested term in the literature.
While it is important to account for the role of power in a way that is fruitful for political economy, there are also limits in the research presented in the book to advance unified theories. To his credit, Holcombe points out that his definitions and analysis does not suggest that other ways of understanding power are problematic. One theoretically productive way to examine power is to explore in depth the structure and influence of elites as they are presented in sociology and political science.

To be more specific, Holcombe argues that public choice theories discussed in the book explain how it is that some people are able to use the political process to gain economic power. Theories developed by sociologists and political scientists explain who benefits from the political bargaining process. Among others, Holcombe cites Charles Wright Mills’ *elite theory* (see, for example, Mills 1956). According to Mills, the elite is composed of individuals in government, corporations, and the military, which together constitute a unified group that works for its own interests. The circulation of members of the elite group ensures that economic and political power is combined. There are, however, pluralist accounts in sociology and political science that dispute elite theories. According to prominent researchers, such as Robert Dahl (1956), power is not concentrated in a unified group but is distributed across a number of individuals and groups. In relation to these theories, it is not clear where exactly Holcombe’s analysis of power stands. For example, in chapters 2 and 4 of the book he appears to accept elite theories regarding the composition of the elite group while he also assumes that economic power in market economies is dispersed among different people that compete for economic resources and political influence. This suggests the need to revisit and refine Holcombe’s account from the vantage point of elite and pluralist accounts.

A third issue with the analysis of this book is Holcombe’s examination of democratic institutions. He argues that the idea that voters determine and influence public policy is an illusion. The main purpose of democratic institutions is to provide a peaceful mechanism for replacing those who have political power. This pessimistic view on the role of democratic institutions invites the following question: What kind of institutions allows people with no power to control those with political power? For Holcombe, the idea that democratic institutions can have significant influence on public policy is questionable. Individual voters are typically uniformed on political issues and choose from the options offered by the elite group. However, one can also examine whether
forms of direct democracy, such as participatory and deliberative democracy, can contribute to the setting and funding of public policy. This alternative proposal is too big to be fully addressed in one book review, and many issues related to accountability, composition, and competence have to be examined in detail. But it is worth mentioning that in practice, citizens’ assemblies have recently discussed questions related to new constitutions (Landemore 2020) and have been used to consider possible policy responses to climate change (Duvic-Paoli 2022). Given this, one could re-examine whether democratic oversight can be also an effective check on the abuse of political power.

Despite these concerns, Holcombe’s book offers a thorough treatment of the concept of power, and it is an important resource for scholars of political economy and public choice theory. Moreover, the material is presented in a way that is accessible to those that do not have a have strong background in the social sciences. The book can also be a valuable reference for graduate and undergraduate courses since it provides the students with a good understanding of some of the theoretical challenges faced by political economists that study the role institutions in different societies.

REFERENCES

Vaios Koliofotis is a lecturer at Eindhoven University of Technology and The Hague University of Applied Sciences. His research focuses on philosophy of science, evolutionary ethics, decision and game theory, and behavioral economics. He has published his research in journals such as Studies in History and Philosophy of Science, Journal of Evolutionary Economics, and History and Philosophy of the Life Sciences.
Contact e-mail: <v.koliofotis@tue.nl>

SOLMU ANTILA  
*Vrije Universiteit Amsterdam*

The injustices of international and global conflict, violations of international codes of human rights and protections, and oppressive government policies are examples of moral wrongs committed by states. When moral wrongs occur, victims and bystanders are correct to seek accountability and assign remedial duties, which, ideally, are based on sound moral justifications. Moral transgressions committed by states present a problem for this standard: states are massive collective agents that have extremely complex organisational structures, bureaucracies, and decision-making procedures that often involve the parliamentary representation of its body of citizens—including its internal conflicts. In the face of such complexity, how can we justly assign and distribute accountability and remedial duties between state officials, government institutions, and citizens? In *Responsible Citizens, Irresponsible States*, Avia Pasternak dives into this moral complexity, and explores whether, how, and under which conditions responsibility to amend states’ wrongdoings should be distributed to its body of citizens specifically.

*Responsible Citizens, Irresponsible States* creates a context-sensitive, practical model for responsibility distribution for states’ wrongdoings. First, it constructs an ambitious argument for the view that, under sufficient conditions of democracy, active citizenship, and practical feasibility, a state’s citizens indeed ought to adopt remedial responsibilities for their states’ wrongdoings equally, regardless of (some) citizens’ disapproval, resistance, or even ignorance of their states’ actions. *Remedial responsibility* here refers to the responsibility to remedy, mitigate, or fix morally undesirable outcomes, regardless of whether one had *causal responsibility* in bringing about the undesirable outcome in question. Second, it reciprocally argues that these remedial responsibilities do not extend to citizens of authoritarian regimes and citizens who are oppressed by their states.

**Author’s Note:** This research is generously supported by the European Research Council with grant ERC-2017-CoG 771074 for the project ‘The Social Epistemology of Argumentation’. I also thank Måns Abrahamson for his helpful comments and suggestions.
Pasternak’s argument that remedial responsibility for democratic states’ wrongs should fall on its citizens can be summarised as follows. Applying the social ontology of Christian List and Philip Pettit (List and Pettit 2011), Pasternak first establishes that states ought to be understood as corporate moral agents, which involve specific kinds of distributions of culpability and responsibility to amend wrongs. Culpability and responsibility to amend moral wrongs can be distributed either in proportion to which government members caused or are blameworthy for the moral wrong, or they can be let to fall to the population at large (chapter 1). Pasternak calls these two competing models the proportional model and the non-proportional model of responsibility distribution, respectively.

Of these two, Pasternak argues for a non-proportional model of responsibility distribution: in the case of moral transgressions committed by states, responsibility for moral transgressions indeed need not fall in proportion to who committed or caused the moral transgression, and often should be let to fall on the state’s body of citizens (chapter 2). An initial motivation for adopting this model comes from the complexity of the collective agent of the state: Pasternak argues that adopting the non-proportional model is prudent because it foregoes the extreme cost and practical infeasibility of tracking and mapping blame for a moral wrong through such a complex collective agent as the state (34, 46).

Cost and feasibility comprise only one factor that can suggest distributing responsibility non-proportionally. Specifically in the context of the state and state wrongdoing, Pasternak argues that the non-proportional distribution model is further strengthened by the fact that in democracies, citizens ‘intentionally’ belong to their states (chapter 3). *Intentional citizenship*, for Pasternak, suggests that citizens genuinely act together with and participate in the state, voluntarily, knowingly, and intentionally accepting their role in the group decision-making apparatus of the state and their role in the maintenance of the state as a cohesive group. This suggests that citizens should also adopt a non-proportional share of remedial responsibilities, regardless of whether or not they agree with or are aware of their states’ actions: for Pasternak, disagreement or lack of awareness do not override the overall intentionality in the practice of participating in the state. The intentional citizenship condition is necessary for non-proportional distribution of responsibility, but empirically contingent; conversely, if the body of citizens overall does not recognise their intentional role in participating in the state, or their participation was
forced or not genuine, citizens would *not* be obligated to adopt remedial responsibilities on their state’s behalf when called upon to do so. As it happens, Pasternak illustrates that surveys of citizens of liberal democratic states (specifically: *Eurobarometer*, years 2005 and 2014; *Afrobarometer*, year 2014/2015; *World Values Survey*, years 2005 to 2014; *International Social Survey Programme National Identity Survey*, years 1995 and 2003; *European Values Study*, year 2010) indicate that citizens feel a strong affective connection to their nations and strongly self-identify with their nationalities. While Pasternak states that these surveys do not directly track intentional citizenship, she claims that they approximate it (97) (chapter 4).

Finally, the third factor that must be taken into account when considering a distribution of responsibility to the population at large is the nature of the state’s regime that caused the moral wrong. Pasternak suggests that in authoritarian states, non-democratic states, and states that oppress their citizens, citizens cannot be expected to adopt remedial responsibilities because they have an inappropriate relationship to the state for such a distribution to be morally justified. In authoritarian states, citizens do not enjoy intentional citizenship because they cannot be said to be inclusive authors of the state’s policies by the very fact that the states are authoritarian. As an example in which moral intuitions about authoritarian regimes and non-proportional distribution of responsibility clash, Pasternak argues that the United Nations Compensation Commission was wrong to make demands of Iraqi citizens in the aftermath of the Gulf War (159–171).

Together, these three factors create Pasternak’s non-proportional responsibility distribution model that can be used to assess whether a responsibility distribution to the population at large is morally justified (chapter 5). In practical terms, Pasternak predicts that this process is likely to lead to the conclusion that, by and large, in democratic states, an equal distribution of remedial responsibilities will be overall justified, while in many authoritarian states it will not be (123). The survey data described in chapter 4 indicates that intentional citizenship is highly correlated with democracy. When an equal distribution among citizens cannot be justified after an assessment of the three factors, alternative solutions should be weighed against the option of imposing the burden on the state’s citizens: leaving the burden with the victims or transferring some of it to the international community (chapter 6).
Lastly, chapter 7 examines how the model applies to the case of historical wrongs and their responsibility-distributive effects. Pasternak argues that in these cases the model still applies; also, in the case of historical wrongs, present day citizens should be expected to adopt non-proportional amounts of responsibility, given the same conditions of state regime change, and in the historical case, state secession (and succession), as in her account. These restrictions have normative implications for international law, which, though it does not think regime changes alleviate responsibility, does not discriminate between regime types.

Pasternak provides a bold, unintuitive argument that is sensitive to and prescriptive for national and international political and legal practice. *Responsible Citizens, Irresponsible Citizens* is laudable for including the perspective of authoritarian states, which are under-discussed as cases in contemporary, liberal democracy-centric analytic political theory. Nonetheless, I show that her account is open to six challenges: two regarding the costliness of blame-tracking, three regarding intentional citizenship, and one regarding a potential negative participation incentive.

The first challenge takes issue with Pasternak’s argument from the costliness of blame-tracking. While it is plausible that blame-tracking can be costly and practically challenging, beyond suggesting that each case should be evaluated individually, Pasternak’s account does not provide principles to guide decisions about when costs are too high, what precisely constitutes practical infeasibility, and why. This challenge does not directly undermine Pasternak’s account, but it suggests that the account would be significantly stronger were more specified conditions and reasons for them provided.

However, a second, more directly threatening challenge related to blame-tracking remains. Even when blame-tracking is extremely costly and practically infeasible, it is not clear how the costliness and infeasibility of blame-tracking help determine a moral shift in responsibility from states to citizens. While even a complete infeasibility (as opposed to practical difficulty) to determine responsible parties would be unfortunate, it neither absolves the responsible parties or automatically suggests moral reasons to create a remedial responsibility for causally non-responsible parties. In fact, if costliness suggests that blame-tracking should generally not be carried out, Pasternak’s argument seems to create a morally vicious incentive for state leaders and actors to create conditions that make blame-tracking as difficult as possible if they recognise this as a path to a kind of moral non-accountability. While neither of these challenges
directly concern the distribution of remedial responsibility if (1) remedial responsibility can be justifiably separated from causal responsibility, (2) individual state officials alone do not have the means to repair the damages they are causally responsible for, and (3) there is a moral justification that the damages should be repaired, they do show that shifting remedial responsibility from state officials to citizens cannot be easily motivated by the costliness of blame-tracking.

If this counterargument holds, Pasternak’s non-proportional model can still be motivated by intentional citizenship. However, Pasternak’s argument from intentional citizenship also faces three challenges. Firstly, despite Pasternak’s claim that citizens voluntarily and intentionally endorse and participate in the collective agent of the state, the survey evidence for the prevalence of intentional citizenship does not obviously suggest that the voluntariness and intentionality conditions of responsibility are fulfilled. Citizens’ self-reporting to self-identify with their state and form an affective attachment to the state does not by default also indicate intentionality or voluntariness to participate in the maintenance and (especially) the decision-making of the state. It is not obvious how an affective or identity endorsement of the state “renders one the inclusive author of what the group does” (71). In social scientific terms, it is not obvious that the operationalisation of intentional citizenship is successful.

Secondly, even granting the intentionality and voluntariness conditions of responsibility, the epistemic or awareness condition of responsibility—which Pasternak discusses under “information” (121)—does not seem to be fulfilled. Pasternak argues that the survey data suggests that citizens recognise their role in executing the group’s decisions, but it is not obvious that it does: once again, the data merely shows that citizens of liberal democratic states generally have a positive affective relationship to their nation and identify with their nationalities. Whether or not they understand that this role imbues them with the rights, duties, and future remedial responsibilities connected to their citizenship is a different question altogether. Again, in social scientific terms, this is a second issue with the operationalisation of intentional citizenship. Of course, this challenge does not hold if the satisfaction of the epistemic or awareness condition could be shown by other empirical evidence.

Thirdly, Pasternak’s model seems to be insensitive to the fact that, like the cost and feasibility of blame-tracking, intentional citizenship is also a matter of degree. Again, even if we accept that intentional citizenship can be gauged accurately and can be morally justified to motivate a
non-proportional responsibility distribution, beyond a suggestion to evaluate each case individually, it is not clear how much intentional citizenship the account requires. And again, this challenge does not directly undermine Pasternak’s account, but only suggests that it would be stronger were more precise principles and conditions regarding the required quantity of intentional citizenship articulated.

Another challenge for Pasternak’s model is that the equality of remedial responsibilities for citizens indicates that citizens will lack any opportunity to decrease their future remedial responsibilities. On Pasternak’s account, even citizens who might accept some level of general remedial responsibility have no possibility to decrease this remedial responsibility even by attempts to prevent the very causal sequence relevant to that responsibility. Consider, for example, an activist who self-identifies as a member of their home state and has an affective attachment to it but who also dedicates a substantial amount of their personal resources to prevent and campaign against a morally disastrous state action. According to Pasternak’s equal membership-based model of responsibility distribution, this activist holds equal responsibility for that action to those citizens who did nothing or helped cause it. Intuitively, while the activist can still adopt responsibility in amending the consequences of the state (and it might be morally commendable for them to do so), they are, at least intuitively, less remedially responsible than others. In fact, the equality of citizens’ remedial responsibilities seems to create a second, parallel vicious incentive for citizens to distance themselves from affairs of the state and towards non-participation. Worryingly for liberal and republican democratic theory, if a citizen chooses to participate in the affairs of the state, she may not only be unrewarded but in fact punished for her intentional citizenship in being obligated to accept responsibility (together with her compatriots) for something that she (or they) had no part in causing. While a proponent of Pasternak’s account might respond that the moral intuitions discussed here are simply faulty, Pasternak’s non-proportional model of responsibility distribution would be strengthened if they could be explained away, which, as the account stands, they are not.

Despite these challenges, the achievement of Responsible Citizens, Irresponsible States should not be unrecognised. Pasternak’s original framing and bold argumentation creates new ground for political philosophy to think about how best to conceive (or even begin to investigate) the political ethics of responsibility given varying feasibility conditions of accountability.
Further work in this field could also consider for example the ethics and philosophy of responsibility in connection to political and historical guilt. For example, the collective, national guilt of the second world war in Germany plausibly affected the kinds of remedial responsibilities Germans were willing to take on. Austria, on the other hand, at least seems to have less historical guilt over the second world war because of a myth of Austria as the first victim of Hitler's regime (see for example Pollak, 2003). Canada is currently dealing with connected questions of guilt and remedial responsibility over their ‘re-education’ of indigenous peoples. Pasternak’s arguments and account could be used in discussion with questions regarding guilt and remedial responsibility: How do collective feelings of guilt connect with remedial responsibilities, intentional citizenship, and oppressive states, including, in the Canadian case, oppression by (seemingly) liberal states? Is guilt an upshot of intentional citizenship, and need it be? In which ways does guilt help or hinder remedial responsibility uptake?

Moreover, Pasternak’s book can also be read as an important contribution in the more general but under-discussed questions for political philosophy about what it actually means for a state to commit a moral wrong and what the consequences for such wrongs might be. When states commit wrongs, what does the collective agent consist of? Following intentional citizenship, to whom can and should we attribute remedial responsibility, and to whom can and should we attribute the actions of the state? This ground is especially fertile and relevant in points of history that are characterised by national and global problems and their consequences in which actions taken by nation states have huge moral implications—such as today’s climate crisis, pandemic, and various international violent conflicts to name only a few.

REFERENCES

Solmu Anttila is a PhD candidate at the Vrije Universiteit Amsterdam. They are currently working on a PhD thesis on the political theory of epistemic injustice.
Contact e-mail: <o.e.anttila@vu.nl>

ERWIN DEKKER  
Mercatus Center, George Mason University

Keith Tribe presents his book, *Constructing Economic Science*, as a continuation of the project to study the institutionalization of political economy initiated by Istvan Hont in the 1980s. That institutionalization entails several related but different ‘institutions’: research journals, academic departments, professional conferences, methodologies, and the training of graduate students, among others. Tribe suggests that we should not ask how a pre-existing economics became institutionalized, but instead: ‘How ‘modern economic thinking’ became embedded in institutional sites, [into] the way in which the creation of the modern university shaped and validated what counted as ‘modern economic thinking’” (9). Therefore, his new book seeks to analyze how different modern university settings in England, which emerged around 1900 shaped economic education, and how this, in turn, influenced the character of modern economics.

To simplify, the motivating problem of the book is how economics 101 became a principles of economics course, and how the fact that it became a principles course has influenced the discipline. I think it is fair to say that Tribe, although he has this ambition, never gets to the second part of this dynamic. He connects his book to other studies of the mathematization of the discipline, but he does not provide an analysis of the way in which the economics curriculum has influenced economics as a discipline. What he does instead is investigate why the contemporary economics curriculum looks the way it does. This itself is a very important question: recent student initiatives, such as the CORE Project and Rethinking Economics, have critiqued and challenged undergraduate economic education, and attempts at reform have in the past frequently ran into institutional barriers.¹ Tribe does not mention these recent conflicts, but he does argue that the version of economics which dominates university education harms students. He suggests that most of them would be better

served by an economic training more closely related to vocational training and practical economic knowledge. He is at pains to demonstrate that, especially in the first half of the twentieth century, the innovations in the economics curriculum developed by Alfred Marshall at Cambridge and Lionel Robbins at the LSE were not developed with the students in mind, but instead used to promote their personal conceptions of what economic knowledge was. His institutional histories of these two universities, which take up the bulk of the book, are contrasted with Oxford to illustrate how a separate discipline of economics and an associated bachelor's degree could only develop when the right academic preconditions were present.

As in his earlier work, Tribe has produced an alternative history which challenges dominant historical narratives. In the introduction he repeats his dislike of histories which reason back from the present to explain how we got here, without considering that things could very well have turned out differently. But this time around it is clear that Tribe is also quite motivated to illustrate the origins of what he perceives to be the ills of the present:

My own account of this “discipline” treats it as an increasingly arcane body of knowledge whose adepts required years of university training in order to be able to speak with any authority on whatever was deemed to fall within their field of expertise. [...] This contradictory quality of economic knowledge—as a special kind of knowledge requiring no legitimation other than a mastery of its techniques, hence self-validating and solipsistic—is a product of the institutional structures within which it is deployed. (18)

His critical perspective is sharpened by the fact that Tribe has worked for years as independent scholar with only tangential relationships to ‘the academy’. The genealogy of the modern economics curriculum that the book provides could thus be read as a belated complement to the infamous surveys that Arjo Klamer and David Colander conducted with economic graduate students at elite institutions which demonstrated that they valued econometric techniques and mathematical skills far more than knowledge and understanding of the ‘real’ economy (Colander and Klamer 1987).

Tribe distinguishes three key features of education: the exam, the content of the courses, and the form of instruction. Let us look at these in turn. The restructuring of the exam, the tripos at Cambridge by Alfred Marshall, is the subject of the second part of the book. During the nineteenth century Cambridge examined its students on the Moral Sciences
Tripos which was a broad exam which included ethics and other social sciences. Marshall managed to convince his contemporaries of the necessity of a separate Tripos for Economics and Political Science which was first implemented in 1905. Tribe situates this within a broader shift within in British academia away from the classical Humboldtian university in which the ideal of Bildung was central and (moral) philosophy enjoyed most prestige, to the modern university model in which disciplines became increasingly separated, and where there was more space for applied sciences such as engineering.

The curriculum which Marshall developed, as part of the new three years honors program in economics at Cambridge, was not yet focused on principles. Economic and general history dominated the program, and although there were three courses on economic principles, students would also take at least two courses each on employment, taxation, and politics and law. Tribe demonstrates that while both Jevons and Marshall wrote an important theoretical treatise they were equally interested in industrial statistics and more applied subjects.

The real revolution in the curriculum, suggests Tribe, came with Lionel Robbins at the LSE. This institution had been a modern ‘university’ from the start and its economics degree was initially tightly integrated with training in commerce. In his earlier work Tribe (1995) has analyzed how Germany pioneered a modern business administration degree around 1900 and in this book he similarly draws attention to commercial degrees which he regards as a more desirable alternative to a narrow economics degree. It is Robbins who sought to break economics free from history, ethics, and commercial education, along the lines he laid out in his programmatic An Essay on the Nature and Significance of Economic Science (1932).

It is not hard to see that Tribe has little sympathy for Robbins’ project. He minimizes the originality of Robbins’ Essay by suggesting that he had only limited knowledge of the German language, although the footnotes of his essay suggest otherwise. The essay is typically read as a bridge between the English political economy and contemporary Austrian economics, but Tribe argues it relies on syntheses of Austrian economics of around the turn of the century. Along the way he lets slip several snarky remarks about the new economics of the Austrians and Lionel Robbins. For instance: “There was a strong tendency by the 1930s for LSE economists to presume that reality ought to conform to theory” (195). Whereas Marshall had still been able to combine his principles with a broad and
extensive knowledge of the real economy, Tribe suggests that this was untrue of those at the LSE. It would take us too far astray to correct this image systematically, but I think that Tribe paints a rather one-sided picture of the LSE here. Hayek, for instance, was hired based on a lecture-series on monetary *history* and one of the most famous products of the LSE in the 1930s was Ronald Coase who throughout his career combined his law and economics with an intimate analysis of real-world problems.

The final feature of education is the form of instruction. On this aspect it is harder to discern a clear historical development in Tribe’s history. He observes that England did not develop graduate education until the final quarter of the twentieth century, which is a significant difference with Germany where it developed a century earlier, as well as the United States which imitated, with about a twenty-five-year delay, the German system. This meant that teaching in England remained oriented toward undergraduates, but how this impacted economics as a discipline is not clearly worked out. In his discussions of Marshall and Robbins he critically evaluates the novelty of their economic work but pays insufficient attention to the fact that both were great synthesizers, and therefore ideal economic communicators, which could also be a reason they thrived in a context of undergraduate teaching. Robbins might not have followed the developments in mathematical economics closely, or read the latest publications in German, but his *Essay* nonetheless became the founding statement of neoclassical economics, and his lectures on the history of economic thought became a classic in the field. What they might have lacked in originality they more than made up for in their breadth of vision and ability to integrate.

Tribe demonstrates, nonetheless, how strong, and distinctive the project of Robbins was. He quotes the LSE professor:

> Surely it is better to push ahead with our analysis, embrace technicality with open arms if technicality will help us, and come to be so frequently right that we acquire the respect now given without question to the practitioners of the natural sciences. (340)

This reinforces the idea that there is a radical break between the knowledge that economics students, or graduates, have, and that of the public, and hence Robbins completely abandons the ideal of a public education or emancipation through economic teaching. It also illustrates that economists suffered from physics envy in more than one way. They did not merely seek to emulate the techniques of physicists, but also
sought the status as an independent field within the modern university, like physics had acquired. Understood this way, it differentiates Robbins’ project significantly from that of Hayek, who not much later started working on his critique of scientism, which criticized precisely the type of ambitions which Robbins had laid out.

Tribe places so much emphasis on Robbins’ ambitions because he is interested in what he calls the ‘scientization’ of economics, a development known in Germany as Verwissenschaftlichung. For Tribe it captures the fact that economics was increasingly cut off from more practical knowledge such as taught in business programs, from historical knowledge, and from ethics. His analysis of the reception of Robbins’ program demonstrates that his contemporaries understood well how radical the demarcation was that Robbins sought to draw.

A key feature of the book is its English focus, and Tribe’s fresh look at the crucial decades around 1900 illustrate the richness of the English historical literature on this subject. Tribe can draw on John Maloney’s work on Marshall and the professionalization of economics, Gerard Koot’s study of the historical economists, Bob Coats’ work on professionalization, Alon Kadish’ oeuvre on the history of Oxford, as well as excellent biographies of the key figures in the book. It provides the book with a lively and polemic character.

But I could not help wonder how the history of economics education in England fitted into broader European and global developments. Chapter nine has an excellent comparison of commercial education in France, Germany, and the United States, later contrasted with Britain. But to what extent the developments toward an insulated economics are mirrored beyond England is unexplored. A few rushed paragraphs toward the end are supposed to make clear how the developments at Cambridge and the LSE paved the way for modern undergraduate teaching in the United States. But Tribe does not connect his work on the quickly expanding literature on the influence on Samuelson, or the different nature of undergraduate and graduate teaching across the pond (Giraud 2018). He claims to be interested in an institutional history but his deep dives into the curriculum and examination at the LSE and Cambridge England are in no way matched by an analysis of how secondary universities in Britain or comparable institutions abroad did or did not copy the English innovations.

Previous studies have suggested that the internationalization of economics education has had the effect of contributing to the focus on prin-
principles rather than institutional and historical differences (Frey and Eichenberger 1993). In the subsequent special issue of Kyklos (Frey and Frey 1995) around the study of Frey and Eichenberger, Alan Peacock even spoke of a professional ‘Gleichschaltung’, in which national differences and idiosyncrasies had to make way for postwar economics U.S. style. But Tribe spends no time on these later institutional changes.

He situates his own work within the context of a broader history of higher education, a subject which he emphasizes repeatedly deserves more attention. It might be for a lack of similar studies, but I would have loved to know how the developments within economics compare to those in other (social) sciences. Inevitably the story that he tells attributes significant agency and novelty to Marshall and Robbins. But if it is indeed true that the proper question is how the modern university shaped disciplinary knowledge, rather than merely a question of how economics was institutionalized, we must know to what extent the developments which Marshall and Robbins initiated were shaped by their environment, rather than their ambitions. Tribe is fond of emphasizing the contingency of the developments and a casual glance at contemporary sociology and political science programs suggest that textbooks do not dominate teaching in these fields in the same way they do in economics, but the book is unfortunately silent on developments in these fields during the same period that economics became a separate science. If the modern university was indeed so crucial in shaping modern economics, would this not be similarly true in neighboring social science fields?

The principal issue with the book I have is, however, that it suggests that Robbins’ project succeeded. Tribe’s argument that the LSE overtook Cambridge as the most influential institution for the development of the discipline ignores a certain Mister Keynes and his disciples who did not only believed to have the upper hand over the LSE in the debate over the causes and solutions for the Great Depression, but who represented what I believe to be a development which became at least as important: the policy-side of modern economics and education. The institutionalization of economics in Britain is most distinct from that in Germany and the United States because it was not driven by policy concerns. The reform-oriented Verein für Sozialpolitik founded in 1873 was from the outset the most important professional association of economists in Germany. Its leading figures such as Gustav von Schmoller sought to transform economics into a historical-empirical science which was directly useful for economic policy. It was imitated in many countries including the United
States where the founders of the American Economic Association were reform-oriented progressives.

A similar organization did not emerge in Great Britain, which is itself an interesting historical contingency. But it should not blind us to the fact that modern economics played a key role in shaping interventionist policy as well as the position of scientific expertise within the state (Dekker 2021). It is true that economics disconnected itself from commerce and business administration, but it reconnected itself to a broader program of social engineering through the state. Tribe suggests in the final paragraph of the book that by the 1970s:

The future upon which Alfred Marshall had set his eyes in 1885—where numbers of young men with “cool heads and warm hearts” might go forth into public and private employment to make good use of their knowledge—had still not arrived. (373–374)

This conveniently ignores the enormous expansion of administrative agencies, policy-oriented centers of economic expertise, and of course real-world interventions that states on both sides of the Iron Curtain engaged in, not just domestically, but also abroad in extensive development programs (Arndt 1987). Economists were not only instrumental in the creation of these institutions, but also found employment in them. These organizations created a steady demand for the new type of economists which graduated from schools such as the LSE, where a lot of modern welfare economics was pioneered and where William Beveridge authored his famous report that laid the groundwork for the modern welfare state. This government demand for economists was partly ‘induced’, but it did much to legitimize the new discipline of economics working on policy questions in the name of ‘the general interest’. If they might serve as ideal types, I would suggest that postwar economics was closer to Keynes’ Cambridge than to Robbins’ LSE.

That does not take away from the originality of the historical perspective that Tribe has brought to bear on the institutionalization of economics and the way it has been taught. The project initiated by Istvan Hont derived its power from the fact that it was designed as a collaborative project which would facilitate international comparisons. Tribe has provided an original analysis of the history of English economics education, which now cries out for similar studies in other countries.
REFERENCES

Erwin Dekker is a Senior Research Fellow with the F. A. Hayek Program for Advanced Study in Philosophy, Politics, and Economics at the Mercatus Center at George Mason University. He has recently published Jan Tinbergen (1903–1994) and the Rise of Economic Expertise (2021) and The Viennese Students of Civilization (2016), as well as the edited volume Governing Markets as Knowledge Commons (2021) all with Cambridge University Press. He is currently working on a project about the descendants of the German Historical School.
Contact e-mail: <edekker@mercatus.gmu.edu>
Many theories of distributive justice deploy thresholds. It is widely believed, for example, that societies have distinct and important duties towards people who do not have enough to meet some minimal standard of living. In contemporary philosophy, this view is called sufficientarianism and was coined by Harry Frankfurt (1987). But thresholds can also indicate that people have too much. Ingrid Robeyns (2017), for example, argues for limitarianism, which is the view that people have too much if they have more wealth than is necessary to live a fully flourishing life. Though not all views in distributive justice can plausibly be characterised as versions of sufficientarianism or limitarianism, many draw on similar thresholds in specifying what justice requires in society’s distribution of benefits and burdens.

In this thesis, I propose an account of the concept and role of thresholds in distributive justice, which is currently lacking in the philosophical literature. I examine what sets threshold views apart from non-threshold views, what exactly qualifies a view as a threshold view, and how threshold views differ from each other. In doing so, I hope to clarify where the conflict between rival distributive views really lies, to shed light on some ongoing misunderstandings about thresholds, and to strengthen the prospects for threshold views in distributive justice.

Following an introduction and overview, in Chapter 2 I propose an account of thresholds in distributive justice (published as Timmer 2021b). I argue that thresholds consist of three elements. First, the level of the threshold determines when people have enough or too much. Second, the moral value of the threshold determines if it is intrinsically valuable for people to reach that threshold or if reaching it is instrumentally valuable. Third, the allocative principles determine the distribution of valuable goods among people above and below the threshold. Among other things, I argue that this account shows that sufficientarians should attach distinct value to maximizing the number of people above the threshold,
which goes against a received wisdom in the literature that such ‘head-counting’ should be rejected (cf. Shields 2012, 103).

In Chapter 3, I propose a novel characterization of sufficientarianism (published as Timmer, forthcoming). I argue that sufficientarianism is best characterised as combining a ‘continuum claim’, a ‘priority claim’, and a ‘deficiency claim’. The priority claim says that we have non-instrumental reasons to prioritize benefits in certain ranges over benefits in other ranges (e.g., a range indicating ‘enough’ and a range indicating ‘not enough’). The continuum claim says that at least two of those ranges are on one continuum. And the deficiency claim says that the lower a range is on a continuum, the more priority it has. Together, these claims say that we have non-instrumental reasons to prioritize benefits below some threshold over benefits above that threshold. I argue that this characterization of sufficientarianism reveals overlooked similarities between sufficientarianism and other views in distributive justice and that it can help strengthen sufficientarianism against common objections, such as the objection that its threshold is arbitrary or has too much normative weight.

Chapters 4–6 focus on limitarianism. In Chapter 4, I defend two versions of limitarianism in distributive justice (published as Timmer 2021a). First, I defend limitarian mid-level principles, which are distributive principles that draw on wealth limits to guide institutional design and individual actions, and which can be endorsed by proponents of different distributive views. For example, both proponents of the idea that justice is ultimately concerned with equality and those who believe justice is concerned with sufficiency can accept that wealth limits promote justice (e.g., by promoting equality or sufficiency), even though they disagree about what justice ultimately requires. Second, I defend presumptive limitarianism, which draws on wealth limits to specify what a just allocation of wealth requires in the absence of substantive grounds to favour specific distributions. Suppose, for example, that wealth should be distributed on the basis of how many hours people work. Substantive principles will then tell us how to distribute wealth if Jane works ten hours and John works two hours. But presumptions tell us how to distribute that wealth if it is not known how many hours people work. I argue that in cases where such presumptive principles are required because there is not enough information to apply substantive principles of distributive justice, justice requires limiting how much wealth people can have.

In Chapter 5, I argue that the prospects for limitarianism depend on both its political feasibility and its likelihood of promoting the relevant
values, such as meeting urgent needs or securing political equality. Limitarians must show that wealth limits are politically feasible and are likely to promote those values. This does, however, raise a challenge because concerns for feasibility and promoting these values might lead to conflicts. For example, higher thresholds may be more feasible whereas lower thresholds may be more effective. I discuss how limitarianism can deal with those conflicts in this chapter, such as by making their wealth limit more directly responsive to concerns for efficiency, or by making them responsive to public opinion about how high, if at all, a justifiable wealth limit should be.

In Chapter 6, I defend limitarianism against Alexandru Volacu and Adelin Costin Dumitru’s (2019) objection that limitarianism is not effective in securing political equality, and that even if it were effective, there would be more efficient means to do so (published as Timmer 2019). I argue that limitarians can support different policies and actions based on their commitment to wealth limits, such as a maximum income and inheritance taxes aimed at dispersing wealth. These policies and actions can be chosen with an eye to efficiency and effectiveness to ensure that limiting how much wealth people can have promotes the values limitarians care about, such as political equality and meeting people’s unmet urgent needs.

Let me end by highlighting two contributions of this thesis for future philosophical theorizing. First, it proposes a shift in the characterisation of theories of distributive justice. Rather than saying, for example, that egalitarianism, prioritarianism, and sufficientarianism, are different distributive views, we should examine them as specifications of a single conceptual core, which often includes a threshold. Aside from offering conceptual clarity, this shift also benefits theorising about eclectic and hybrid distributive views, which include concerns for various distributive concerns, such as equality, priority, and sufficiency. Second, the account of thresholds in distributive justice can be developed into an account of thresholds in philosophy more generally. Though developing such an account is beyond this thesis’ scope, it would be a valuable contribution to contemporary philosophical theorising.

REFERENCES

**Dick Timmer** is an Assistant Professor at TU Dortmund University. He works in political philosophy and ethics, with a particular focus on distributive justice. His work has been published in journals such as *Journal of Applied Philosophy*, *Journal of Political Philosophy*, *Philosophy Compass*, and *Utilitas*.
Contact e-mail: <dick.timmer@tu-dortmund.de>
PHD THESIS SUMMARY: Desert, Luck, and Justice

HUUB BROUWER
PhD in Philosophy, January 2020
Tilburg University

Nearly 2400 years ago, Aristotle wrote that “all people agree that what is just in distribution must be in accord with some sense of desert” (Nicomachean Ethics 1131a). Since then, many philosophers have written approvingly about desert as a principle of distributive justice. Also, a host of empirical studies claim to show that people are—as Shelly Kagan (2012, 3) puts it—"friends of desert": they think that some people are more deserving than others, and that it is a good thing if people get what they deserve (see Konow and Schwettmann 2016 for an overview).

Nevertheless, the popularity of desert as a principle of distributive justice waned during the 1970s, coinciding with the publication of John Rawls' (1971) desert-less theory. My PhD is a collection of six chapters that all aim to contribute to answering the question: What role, if any, can desert plausibly play within egalitarian theories of distributive justice?

Ongoing philosophical reflection on the notion of desert has made clear that it is surprisingly complex: there are many different conceptualizations of it, and these are not all vulnerable to the same objections. I argue that, because of this complexity, generalized skepticism about desert as a principle of distributive justice can and should be resisted.

Chapter 1 of my PhD is introductory. I explain that it remains very controversial what ‘desert’ is, exactly. Many philosophers agree that desert claims are three-place relations, uniting a desert subject (S), a desert object (O), and a desert base (B). An example would be the claim that Rosemary (S) deserves to win the Rotterdam weightlifting contest (O), because she has the strongest muscles of all contestants (B). However, there is considerable debate about what permissible desert subjects, objects, and bases are and about when these stand in an appropriate relation to each other. In this debate, three ‘received wisdoms’ have emerged:

Received Wisdom 1 (aboutness principle): The only permissible B are acts and attributes of S (cf. Feinberg 1970, chap. 4; Olsaretti 2003).
Received Wisdom 2 (responsibility requirement): When S deserves O because of B, then S is responsible for B (cf. Feldman 1995; McLeod 1998).

Received Wisdom 3 (backward-lookingness): If at time t, S deserves O because S performed action or had a certain attribute B at t’, then t’ cannot be later than t (cf. Feinberg 1970, chap. 4; Feldman 1995).

In my thesis, I challenge Received Wisdoms 2 and 3. I argue that rejecting these opens avenues for desert to play a plausible, but limited role in theories of distributive justice.

Chapters 2 and 3 of my PhD are conceptual, contributing to debates about what desert is. Chapter 2 examines whether it is defensible to hold that desert should play a (very) different role in distributive and in retributive justice. John Rawls (1971) certainly thought so, but did not provide a satisfactory defense of the asymmetry of desert in his own work (Honig 1993, chap. 5). Samuel Scheffler (2000), Saul Smilansky (1996), and Jeffrey Moriarty (2003, 2013) have all proposed defenses of asymmetries of desert. Their defenses, however, have either been convincingly criticized or only hold for weak asymmetries of desert—in which desert plays a role in both retributive and distributive justice, but a (somewhat) more central role in the former. These defenses do not support strong asymmetries, such as the asymmetry present in Rawls’ work, in which desert plays no role at all in distributive justice, but a central role in retributive justice.

I defend two main claims about strong asymmetries of desert. First, I argue that some recent defenders of desert-based theories of distributive justice may actually be interested in defending asymmetries that go the other way: in which desert plays a more central role in distributive justice rather than in retributive justice. Second, I argue that strong asymmetries in both directions can be defended. The responsibility defense is premised on the thought that a rejection of Received Wisdom 2 (responsibility requirement) for desert claims on certain objects of distributive justice can be defensible, whereas this is not the case for desert claims on the objects of retributive justice. The thought motivating the shameful revelation defense is that implementing desert-based theories of justice would require people to reveal facts about themselves that they consider shameful, and that forcing them to do so would, under certain conditions, be permissible in the retributive sphere, but not in the distributive sphere. I conclude that there is no reason for those who are sympathetic to strong asymmetries of desert to revise their thinking radically.
In chapter 3, I probe *Received Wisdom 3*: that desert claims are always backward-looking. Fred Feldman (1995) was the first to challenge this received wisdom: he argues that people can deserve compensation on the basis of their future acts and suffering. There have been two main responses to Feldman’s challenge. The first is to resist it and argue that desert really is a backward-looking notion (Celello 2009). The second is to build on Feldman’s suggestion that desert may sometimes be forward-looking and propose a more elaborate account of forward-looking desert (Schmidtz 2002). I argue that desert is forward-looking for a wider range of objects and bases than has been considered so far. To make this claim, I proceed in two steps. First, I introduce Feldman’s argument for the forward-looking desert of compensation and argue that criticisms that have been raised against it can be resisted. Secondly, I argue that not just the desert of compensation, but also chances, praise, and rewards can, sometimes, be forward-looking.

In chapters 4 and 5, I move along to the relation between desertism and luck egalitarianism. Luck egalitarianism is, very roughly, the view that “inequalities in the advantages that people enjoy are acceptable if they derive from the choices that people have voluntarily made [option luck], but [...] inequalities deriving from unchosen features of people’s circumstances are unjust [brute luck]” (Scheffler 2003, 5). Many luck egalitarians invoke desert, and it is not always clear when the two views differ. In chapter 4 (co-authored with Thomas Mulligan), I ask what the differences are between desertism and luck egalitarianism. I argue that desert and luck egalitarianism come apart in three important contexts. First, compared to desertism, luck egalitarianism is sometimes stingier: it fails to justly compensate people for their socially valuable contributions—when those contributions arose from option luck. Second, luck egalitarianism is sometimes more restrictive than desertism: it fails to justly compensate people who make a social contribution when that contribution arose from brute luck. Third, luck egalitarianism is more limited in scope: it cannot diagnose economic injustice arising independently of comparative levels of justice.

In chapter 5 (co-authored with Willem van der Deijl), I ask whether desert could offer a plausible principle of stakes for luck egalitarianism. A principle of stakes specifies “what consequences can justifiably be attached to the features that are the appropriate grounds of responsibility” (Olsaretti 2009, 167). Without it, luck egalitarianism is indeterminate: the theory cannot tell us what consequences people should bear when they
are responsible for their acts. Olsaretti (2009) argues that desert cannot serve as a plausible principle of stakes for luck egalitarianism. In this chapter, I defend the claim that she is too pessimistic about this by introducing a simple, but plausible, desert-based account of stakes that is immune to her argument.

Chapter 6 (co-authored with Dick Timmer) takes a more applied perspective by focusing on desert and high salaries. Specifically, it asks whether it is possible to earn too much. I observe that in the philosophical debate about income and wealth, one assumption is surprisingly uncontested. This is the assumption that, all else equal, people should always have the freedom to earn more income and accumulate more wealth. I argue that this assumption should be rejected. In particular, I argue that three prominent approaches to taxation and justice in political philosophy—Rawlsian egalitarianism, utilitarianism, and desertism—all hold that above some point, people do not have a justice-relevant, non-instrumental claim to be free to earn more income. I then discuss instrumental arguments against income limits and argue that those arguments are weaker than is often thought. I conclude that it may, in fact, be possible to earn too much.

In 1971, John Kleinig remarked that “the notion of desert seems by and large to have been consigned to the philosophical scrap heap” (71). In my thesis, I hope to have contributed to showing that philosophers should not be too quick in consigning desert to the scrap heap of distributive ideals. Several received wisdoms about it can and should be challenged: the connection between desert and responsibility, on the one hand, and desert and time, on the other, are more complicated than it seemed. The rejection of received wisdoms about desert opens new and exciting avenues for philosophical research into desert as a principle of distributive justice.

REFERENCES


**Huub Brouwer** is an assistant professor of Ethics and Political Philosophy at Tilburg University. Huub has been on research visits to Gothenburg University, Oxford, Pompeu Fabra University, and Yale. He mainly works on topics in distributive justice—particularly on desert, property, and responsibility-sensitive egalitarianism.

Contact e-mail: <h.m.brouwer@tilburguniversity.edu>

Website: <huubbrouwer.com>
PHD THESIS SUMMARY: 
Choice Architecture: A Message and Environment Perspective 

LUCA CONGIU
PhD in Economics, February 2020 
University of Insubria

Contrary to standard theories of decision making, which posit that preferences are invariant across logically equivalent choice scenarios, experimental evidence shows that even theoretically irrelevant aspects of a decisional context can affect choices (for an overview, see Shafir 2013, especially Part 8). For instance, food choice may be affected by the prominence or order of items on display, whereby those that are more prominently displayed may draw the consumer's attention and be selected in response (e.g., Dayan and Bar-Hillel 2011). The idea that supposedly irrelevant features of the choice context matter for the final decision is encapsulated in the concept of 'choice architecture', that is, the way in which a choice situation is 'designed' and presented to the decision maker (Thaler and Sunstein 2009; Thaler, Sunstein, and Balz 2013). More specifically, a choice architecture includes not only the options available and the budget constraint, but also the options' order and prominence, the way in which they are described to the decision maker (e.g., if information is framed in positive or negative terms), the presence of default options, the rules or norms that govern the decision, and so on. Given the plethora of aspects that form a choice architecture (and that can affect the final decision), decision theorists and policy makers have strived to advance models and frameworks to analyze and design choice architectures (e.g., Johnson et al. 2012; Thaler, Sunstein, and Balz 2013; Lades and Delaney 2022).

In my doctoral thesis, I join this endeavor by advancing a framework which conceptually subdivides choice architecture into two dimensions: 'Message' and 'Environment'. Roughly speaking, 'Message' refers to the verbal description of options (both oral and written) and information or suggestions about the choice at hand, while 'Environment' refers to every element that the decision maker can find in her physical or virtual surroundings, perceive through her senses, and interact with. The core idea is that a choice situation may be designed by working on these two dimensions and that choice behavior depends on how either dimension is
designed in relation to the other. Ultimately, I argue that analyzing and designing choice architectures through the Message-Environment distinction allows for a fruitful, novel perspective on choice behavior and policy making. This idea is laid out in three chapters.

Chapter 1 (published as Congiu and Moscati 2022) reviews the literature on choice architecture, with particular attention to so-called ‘nudges’ (Thaler and Sunstein 2009). Nudges are broadly defined as aspects of the choice architecture that steer people’s behavior in welfare-improving directions by acting on their cognitive limitations and biases (Thaler and Sunstein 2009; Hansen 2016). An intense debate has flourished around the concept of nudge, focusing on three main themes: (1) the exact definition of a nudge; (2) the justification for the use of nudges in policy making; and (3) the effectiveness of policies based on nudges. On the first theme, the chapter argues that, based on the existing definitions of nudge, it is not always straightforward to separate nudges from other tools for behavior change, such as standard policies (e.g., taxes and other monetary incentives) and marketing techniques (e.g., prominently displaying expensive products). With respect to the justification issue, the chapter clarifies that ‘pro-self’ nudges, that is, those that aim to increase the nudged person’s welfare, can be adequately justified by ‘libertarian paternalism’—an approach to policy making that authorizes steering people’s behavior in directions that improve their welfare while preserving choice freedom (Thaler and Sunstein 2009). By contrast, ‘pro-social’ nudges, which aim to increase society’s welfare (sometimes at the expense of the nudged person), are not adequately justified by libertarian paternalism and call for alternative justifications for their use (e.g., approval by the public). Finally, the chapter shows that nudge policies can be as effective as standard policies (and in some cases even more effective), although the two should be regarded as complements rather than substitutes.

The Message-Environment framework is introduced in chapter 2 (published as Congiu and Moscati 2020). As anticipated, the framework decomposes choice architecture into two dimensions: Message and Environment. The Message dimension includes all verbal communications, both oral and written, that describe the choice situation or provide some information about it. Examples are messages such as ‘Smoking damages the lungs’ that attempt to influence people’s behavior by targeting their motivation to preserve their well-being, and ‘Suggested donation: £10’ which
aims at establishing a reference level or ‘anchor’ for donations. The Environment dimension includes all elements of the choice context that the decision maker can encounter in her physical or virtual surroundings, perceive through her senses, and interact with. Among others, the Environment includes the way the options are ordered or displayed (e.g., items on a shelf) and the presence of constraints (e.g., budget constraints), obstacles (e.g., speed bumps), or default options (e.g., double-sided printing). The framework is then applied to analyze some common types of nudges, and specifically to identify a Message and an Environment counterpart for each. For example, it is argued that anchors can be provided at Environment, such as when donors are provided with a list of potential contributions (e.g., \{€1, €2, €5, …\}), but also at Message, through communications such as ‘Limit of 10 per person’ or ‘How many units do you usually buy?’ Finally, the Message-Environment framework is applied to the analysis of Amazon’s website (i.e., a digital choice architecture) and to the design of a fictitious choice architecture to foster saving for retirement.

Chapter 3 (published as Congiu 2022) applies the distinction Message-Environment to the architecture of a choice involving risk, namely, an elicitation task for risk aversion. Typically, risk aversion is elicited experimentally by presenting the decision maker with a menu of monetary lotteries with varying risk and asking her to choose the lotteries she prefers. On their part, lotteries can be presented through a verbal description stating the outcomes and their likelihood (e.g., ‘Win $5 with probability 10%', ‘1/10 chance to win $5')—that is, they have a Message dimension—and can be accompanied by a pictorial display, such as a pie chart or bar graph—that is, they have an Environment dimension. The literature on risk communication suggests that alternative but logically equivalent numeric formats (e.g., percentages vs. ratios) and pictorial displays (e.g., continuous vs. discrete graphs) may frame risk in a way that alters the perception of it (e.g., Schapira, Nattinger, and McHorney 2001). Thus, I design a multiple price list task (Holt and Laury 2002) where risk information is presented at Message as percentages ('10%') or ratios ('1 out of 10') and at Environment as a pie chart sliced either in two or ten slices. Results show that neither the Message framing (adopting ratios) nor the Environment framing (slicing pies) significantly altered the average elicited risk aversion. However, the pictorial framing significantly reduced the elicited risk aversion of those participants who focused on the probability of the high outcome in their decisions, suggesting that the impact
of the pictorial framing may depend on which probability in a choice under risk draws the most attention.

In conclusion, the Message-Environment framework has as its core the idea that choice architecture can be decomposed into a Message and an Environment dimension, and that choice behavior depends on the way in which these dimensions are designed with respect to one another. An important implication is that an intervention in the choice architecture can be in principle more effective if Message and Environment are designed in a complementary way. For instance, a message discouraging the adoption of a certain conduct—e.g., 'Free-riding damages the group output'—can be more effective if combined with an obstacle or constraint to the adoption of that conduct—e.g., a system that detects and punishes free-riding with a given probability. Likewise, an intervention at Environment might be less effective without a complementary one at Message, because without proper communication decision makers might not understand the importance of exhibiting a particular behavior. To continue the example provided above, group members might be more prone to avoid free-riding if they understand the implications of their actions for the group output rather than just focus on avoiding being caught by the system. Ultimately, the Message-Environment framework prompts the ‘architect’ of a choice situation to investigate potential complementarities between the two dimensions, generating in turn more effective interventions.

REFERENCES
Schapira, Marilyn M., Ann B. Nattinger, and Coleeen A. McHorney. 2001. “Frequency or


**Luca Congiu** is Assistant Professor of Economic Policy at Tor Vergata University of Rome. His main research fields are Behavioral Economics and Behavioral Public Policy, with a focus on the effect of choice architecture on individual decisions. More generally, he is interested in the application of behavioral and psychological theories both to explain and affect behavior. Contact e-mail: <luca.congiu@uniroma2.it>

Website: <sites.google.com/view/luca-congiu>
Giving people equal opportunities is a principle of justice that is built on two fundamental ideas. On the one hand, outcome differences across individuals are unacceptable if they are rooted in factors that are beyond individual control. Examples of such circumstance characteristics are the biological sex, race, and socio-economic status of one's parents. On the other hand, if individual outcomes were the result of effort, proponents of an equal-opportunity ethic would accept outcome differences across individuals as fair.

These principles are a core reference point in the philosophical discourse on distributive justice (Cohen 1989; Roemer 1998; Arneson 1989) and they are widely referenced by public and political actors when discussing inequality in various domains of life including health, education, and income.

In this thesis I advance the economic literature on equality of opportunity in three dimensions: i) the construction of measures, ii) the estimation of measures, and iii) the identification of causal drivers.

Chapter 1 is dedicated to the development of inequality measures that combine opportunity-egalitarian principles with other principles of fairness in joint indicators of unfair inequality. It is motivated by empirical evidence showing that people do not judge inequality as problematic per se but that they take the underlying sources of income differences into account (Adriaans et al. 2020). In contrast to this evidence, standard measures of inequality do not adequately reflect these normative preferences. Yet, it is important to take account of these normative preferences if we want to address the widespread perception of unfairness that has stimulated social tension and weakened support for existing economic systems in many countries around the world.

In this chapter, which is based on Hufe, Kanbur, and Peichl (forthcoming), we propose an alternative way of measuring inequality that corresponds more strongly to general principles of justice and the normative
preferences upheld by the larger public. In particular, the proposed measures acknowledge that equality of opportunity is important but individually insufficient to define a fair distribution of resources. For example, many people would subscribe to the moral imperative of addressing extreme outcomes like hunger, homelessness, and material deprivation regardless of how these outcomes came about (Konow and Schwettmann 2016; Cappelen et al. 2013). However, such a preference stands in contrast to the opportunity-egalitarian doctrine according to which we should accept outcomes if they were the result of individual responsibility and effort exertion. In response, we propose the first family of measures for unfair inequality that incorporate the principles of equality of opportunity and freedom from poverty in a co-equal fashion. We, therefore, take seriously the idea that equity is not represented by the absence of any inequality in outcomes, but that it requires life success to be determined by factors outside of an individual’s control and that everybody should have enough to make ends meet.

Furthermore, we provide two empirical applications of our measure that yield important insights for the inequality debate and the design of appropriate policy responses. These empirical applications use data from the Panel Study of Income Dynamics (PSID) and the European Union Statistics on Income and Living Conditions (EU-SILC). Both data sources are household surveys that contain the necessary information to construct our measures of unfair inequality: individual incomes and personal background characteristics that qualify as circumstance characteristics. In particular, we use the following set of circumstances: race and migration background, parental occupation, parental education, and biological sex. First, we analyze the development of inequality in the US over the period 1969–2014 from a normative perspective. Our results show that the US trend in unfair inequality has mirrored the marked increase in total inequality since the beginning of the 1980s. However, beginning with the 1990s, unfair inequality followed a steeper growth curve than total inequality. We illustrate that this trend is mainly driven by a less equal distribution of opportunities across people that face different circumstances beyond their control. Second, we provide a corresponding international comparison between the US and 31 European countries in 2010. We find that unfairness in the US shows a remarkably different structure than in societies with comparable levels of unfairness in Europe. Our evidence suggests that inequality in the most unfair European societies is largely driven by poverty increases that followed the financial crisis of 2008. On
the contrary, unfairness in the US is driven by marked decreases in social mobility.

Chapter 2 is dedicated to the estimation of inequality of opportunity measures. Measures of inequality of opportunity quantify the extent to which individual outcomes are predicted by circumstance characteristics. This idea is commonly operationalized by using a set of circumstances to predict an outcome of interest and calculating inequality in the distribution of predicted outcomes: the more predicted outcomes diverge, the more circumstances beyond individual control influence outcomes, and the more inequality of opportunity there is. However, in standard practice researchers are left to their own devices in specifying the prediction function, i.e., they have to decide which circumstances to include in the empirical model and they have to assume how these circumstances interact. This leads to downward biases in inequality of opportunity estimates if the prediction function is too restrictive to capture the dependence of life outcomes on circumstance characteristics. On the contrary, it leads to upward biases if an overly flexible prediction function overfits the data. Overfitting occurs if the model is too complex for a given sample size. As a consequence of overfitting, the relevant parameters are noisily estimated, i.e., they have very large standard errors which in turn inflate inequality of opportunity estimates (Brunori, Peragine, and Serlenga 2019).

In this chapter, which is based on Brunori, Hufe, and Mahler (2022), we propose the use of machine learning methods—and regression trees and forests in particular—to overcome the issue of ad-hoc model selection. Machine learning methods use algorithms to choose the best estimation models. Therefore, they let the data speak and are not subject to the discretionary choices of researchers. Furthermore, they allow for flexible models of how unequal opportunities come about while imposing statistical discipline through criteria of out-of-sample replicability. These features serve to establish inequality of opportunity estimates that are less prone to upward or downward bias.

To showcase the advantages of machine learning methods we compare them to existing estimation approaches in a cross-sectional dataset of 31 European countries. We demonstrate that current estimation approaches overfit (underfit) the data, which in turn leads to upward (downward) biased estimates of inequality of opportunity. These biases are sizable. For example, some standard methods overestimate inequality of opportunity in Scandinavian countries by close to 300%, whereas they underestimate the extent of inequality of opportunity in Germany by more
than 40%. Hence, cross-country comparisons based on standard estimation approaches yield misleading recommendations concerning the need for policy intervention in different societies.

Chapter 3 is dedicated to the identification of causal factors that drive the existence of unequal opportunities. In particular, I focus on the labor market opportunities for men and women. In many societies, gender gaps in labor market outcomes have closed significantly post-World War II. In response to changing economic incentives, heterosexual couples with children have adjusted their time-use and spending patterns, henceforth leading to marked changes in the way they invest in the skill formation of their children. These empirical trends raise the question of whether opportunity-equalization in one generation, i.e., the closing of gender gaps, has led to a dis-equalization of opportunities in the next generation, i.e., by increasing skill gaps. Skill gaps among children constitute an early indicator of unequal opportunities as they are highly predictive of important life outcomes during adulthood including income, education, and health.

In this chapter, I study how changes in the parental wage gap influence children’s formation of socio-emotional skills as measured by the Big Five personality inventory. I investigate this question by constructing a sample of circa 6,000 German siblings aged 2–12 for whom I observe measures of the Big Five inventory at the same age but in different calendar years. In addition, I match this sibling sample to measures of potential wages that reflect variation in the sex- and education-specific labor demand across commuting zones in Germany. As a result, I can analyze within-family changes in time-use and monetary resources that follow from plausibly exogenous changes in the relative labor market incentives for mothers and fathers, and how these changes affect the socio-emotional development of their children. Importantly, this research design allows for a causal interpretation. First, the within-family comparison rules out all between-family differences including socioeconomic status or region of residence as confounding factors. Second, the use of potential wages addresses concerns of reverse causality that may arise if the development of children co-determined the labor market decisions of their parents.

I find that decreases in the parental wage gap lead to i) an increase in households’ total financial resources, ii) an increase in financial resources controlled by mothers, and iii) an increase in the use of informal care providers. Despite these changes, I find no effect on the socio-emotional
development of children as measured by the Big Five inventory. These null effects are precise enough to exclude various effect sizes from other quasi-experimental interventions studied in the existing literature. In sum, these findings suggest that strides towards gender equality in the labor market do not necessarily come at the cost of detrimental effects on child development.

REFERENCES

Paul Hufe is an Assistant Professor at the University of Bristol. His research interests lie at the intersection of public, labor, and normative economics. In particular, he works on the measurement of unfair inequality and analyzes which factors cause an unequal distribution of opportunities in societies today. Paul Hufe received his Ph.D. from LMU Munich in 2021.
Contact e-mail: <paul.hufe@bristol.ac.uk>
PHD THESIS SUMMARY:
Issues on the Measurement of Opportunity Inequality

HUGO DEL VALLE-INCLÁN CRUCES
PhD in Economics, February 2021
University of Vigo

Recent years have seen a surge in the literature on inequality of opportunity (henceforth, IOp). This is possibly due to its compelling notion of fairness, which resonates in economic, political, and journalistic circles (Bourguignon, Ferreira, and Walton 2007; Manuelyan Atinc et al. 2005). However, the empirical application of IOp is inconsistent and must improve in order to become policy-relevant. The shortcomings of IOp measurement include a lack of correspondence with its theoretical principles, stringent data limitations, and the absence of a standardized, widely accepted methodology. In my dissertation, I have explored these issues.

In this summary, I will first briefly introduce the notion of IOp, then review the contributions of my doctoral thesis. This notion traces back to a debate in political philosophy during the 1970s and 1980s, which shifted the focus of the egalitarian project from ‘equality of outcomes’ to ‘equality of chances’. Building on the work of John Rawls (1971), philosophers such as Richard Arneson (1989), Gerald Cohen (1989), Ronald Dworkin (1981a, 1981b), and Amartya Sen (1980) sought to define a concept of equity that would align individual rewards with personal responsibility. Where political philosophers started, economists soon followed. The seminal contributions of Marc Fleurbaey (1995), John Roemer (1993, 1998), and Dirk Van de gaer (1993) modelled this ideal of fairness and combined it with distributional analysis, systematically analysing which allocations should be deemed ‘fair’ or ‘unfair’.

In short, inequalities in any outcome—be it income, wealth, health status, and so on—are ‘fair’ or ‘unfair’ depending on what they stem from. Inequality caused by factors under an individual’s control—like the degree of effort exerted—is considered fair, while unfair inequality arises from circumstances beyond individuals’ control—like gender or race.
Hence, the goal of the egalitarian project is to create a ‘level-playing field’, which eradicates unfair but preserves fair sources of inequality.¹

My thesis is structured in six chapters. I begin by introducing the philosophical and empirical basis of IOp in chapters 1 and 2, respectively. Chapters 3 and 4 attempt to tackle the abovementioned shortcomings of IOp measurement. Chapter 3 looks at an inconsistency of IOp measurement with respect to its theoretical principles—in particular, that the level of analysis may entail a normative choice regarding gender; and chapter 4 addresses the problem of data limitations in the IOp approach. In chapter 5, building on the contributions of the previous two chapters, I show how to measure IOp in Europe with an unprecedented level of detail. Chapter 6 concludes. In the rest of this summary, I will go over chapters 3, 4, and 5 in more detail.

Chapter 3 tackles an inconsistency of IOp measurement with respect to its theoretical principles. In particular, I scrutinise the appropriate level of analysis, and argue that, when measuring IOp, the level of analysis may entail a normative choice regarding gender. Standard analysis of economic inequality looks at the distribution of income or wealth at the household level. This perspective is justified by the object of study—that is, the access to economic resources that individuals have (Jenkins and Van Kerm 2009)—which may be better captured by taking the household, rather than the individual, as the unit of analysis.² However, the IOp approach is not utilitarian, and the object of study is not the access of individuals to economic resources. Hence, such justification offers no hint to what the appropriate level of analysis should be. As a matter of fact, it is common to see empirical applications of IOp using either the household or the individual level as if they were interchangeable, but this issue carries major effects. The choice entails normative implications and strongly influences IOp estimates.

This chapter begins by showing that IOp estimates are strongly sensitive to the level of analysis. For this I employ one measurement technique (Ferreira and Gignoux 2011) with different outcomes of interest (some at the household level and other focused on individuals). With data from 31

¹ For a survey on the philosophical grounds of the IOp approach, see Ferreira and Pergine (2016), or Roemer and Trannoy (2016).
² As long as we assume (perfect) within-household redistribution, even though this rarely holds. Problems associated with this assumption are well known: see Haddad and Kanbur (1990) for an early reference. Recent contributions include Lechene, Pendakur, and Wolf (2019), who show that poor people may live in non-poor households, Fremeaux and Leturcq (2020), who find stark differences between household and personal wealth inequality, and Sauer, Rehm, and Mader (2021), who offer a general perspective.
European countries—employing the well-known and researched European Union Statistics on Income and Living Conditions (EU-SILC) database—I show that IOp estimates at the household and individual level are, in fact, uncorrelated, and that the latter tend to be much higher than the former. Then, I present a set of theoretical remarks detailing the causes for this discrepancy, which in great part is due to the role of gender in determining IOp. Namely, considering an outcome at the household level implies assuming within-household redistribution of opportunities, which virtually nullifies the importance of the circumstance gender, since most sampled households are composed of both men and women. Notice that pooling all resources of the household, that is, assuming within-household redistribution, means imputing the same amount of outcome to men and women who live together, blurring possible personal differences. Therefore, to be able to capture the effect of gender on IOp the focus ought to be on individuals—if one were to focus on households, the contribution of gender to the overall level of IOp would be obscured, biasing the estimates downwards. Finally, I check this claim empirically with a decomposition technique of overall IOp that confirms gender as a major component when the level of analysis is individuals, and barely relevant when it is households. I conclude by arguing that if one believes gender to be a potential source of IOp, then the appropriate level of analysis should be the individual.

Chapter 4 addresses the problem of data limitations in the IOp approach. In it, I propose a new measurement strategy that relaxes a ubiquitous data constrain, by circumventing the restriction imposed by the scarcity of a key piece of information: the family background of individuals. Family background is widely believed to play a major role in the determination of IOp, and hence, it is normally accounted for when measuring it. The standard proxies for this dimension of IOp are parental education and/or parental occupation, and, although these are naturally imperfect proxies, it is generally assumed that they serve the purpose well. In this chapter, I do not question their accuracy; rather, I deal with their availability. Since data on parental education and parental occupation is scarce, I propose using an alternative proxy, capital income, that is both related to family background and widely available.

---

3 Indeed, this kind of information is scarce: consider the following two well-known databases for the study of poverty and inequality. As of September 2021, the Luxembourg Income Study had information on parental education—parental occupation not available—in 92 of its 497 datasets (around 19%). Likewise, the EU-SILC database had information on parental education and occupation in 3 of its 16 waves (also around 19%).
Using capital income as a proxy for socioeconomic status has a practical advantage—it allows obtaining many new IOp estimates that would otherwise be impossible to obtain. But is it fit for purpose? The relationship between capital income and family background has been explored in a large body of literature studying wealth inequality, the intergenerational transmission of advantages, and the determinants of financial returns. In sum, the measurement strategy I propose works by leveraging the intergenerational persistence of wealth.

Given these theoretical foundations, I then proceed to the empirical validation. For this I also use data from the EU-SILC database, testing the accuracy of the method in 31 European countries with satisfactory results. The validation method consists of comparing standard estimates (obtained with a well-established methodology in the field, that is, using parental education and occupation to account for family background) with the ones obtained with the measurement strategy I propose. To the extent the latter are similar to the former, I conclude that the strategy is reliable. The differences reside, naturally, in the number of estimates I can obtain, not in their magnitude. Notice that I can only compare a few of the data points produced with my method, since they are many more than the ones generated with a standard methodology—the assumption is that if those data points that can be compared perform reliably, the rest will do so as well. In sum, I obtain very high correlations between the two sets of estimates. The correlations are close to one, both rank and pairwise, and the magnitudes are very close as well.

The fifth, and final, chapter demonstrates the benefits of the new measurement strategy by estimating and tracing the evolution of IOp in Europe during and after the Great Recession (2004–2017). The core contribution of these new IOp estimates is their unprecedented level of detail, which is only possible due to the use of capital income as an alternative proxy for family background. The main takeaway from these data is that, contrary to what might have been expected, IOp did not increase in most European countries during the Great Recession. Furthermore, no clear pattern arises across the continent; some countries saw their IOp increase while others saw it decrease. Even more puzzling, no relationship appears to exist between the extent to which an economy contracted during the Recession and the evolution of IOp its inhabitants endured.

4 In the thesis, I devote a whole section to a review of this literature. Prominent references include Alvaredo, Garbinti, and Piketty (2017), Charles and Hurst (2003), and Von Gaudecker (2015).
A second contribution of this chapter is a sensitivity analysis. More precisely, I use the EU-SILC panel data to assess the sensitivity of my IOp estimates to short-term income shocks. It is well-known that inequality estimates based on cross-sectional survey data might be distorted by transitory shocks to the income that individuals report. To assess the extent to which my estimates suffer from this problem I proceed to estimate IOp following the same individuals across four consecutive years. I find no support that my IOp estimates, obtained using the EU-SILC panel data, are sensitive to short-term income shocks. What is important to stress here is the following: since the EU-SILC panel data does not contain information on parental education, carrying out this robustness test is only possible thanks to using capital income as a proxy for family background. Without such an alternative proxy for family background, it would not have been possible to check, using the EU-SILC database, whether short-term income shocks may affect IOp estimates.

The IOp approach is very promising for the advancement of the egalitarian agenda. Still, it faces many methodological problems that prevent it from becoming policy-relevant. My thesis is an attempt to tackle two of these issues, one theoretical and the other empirical. First, I argue that in order to account for the effect of gender, estimates of IOp must take the individual, rather than the household, as their unit of analysis. Second, developing informed policies that target morally relevant inequality is severely limited by the lack of information on individuals. My thesis relaxes this limitation with respect to family background, an important determinant of IOp, by proposing—and motivating, both theoretically and empirically—capital income as an appropriate proxy for this dimension.

REFERENCES

**Hugo del Valle-Inclán Cruces** obtained his PhD from the University of Vigo, where he worked as researcher and teacher. He is a member of ECOBAS. His dissertation is available in del Valle-Inclán Cruces (2021). Contact e-mail: <hvalle-inclan@uvigo.gal>
PHD THESIS SUMMARY:
Three Economic Extensions of John Rawls's Social Contract Theory: European Union, Tax Compliance, and Climate Change

KLAUDIJO KLASER
PhD in Economics and Management, June 2019
University of Trento

The overarching framework of my thesis concerns the relationship between ethics and economics (Sen 1987), under the idea that ethics can inform economic theory and practice in various ways (White 2019). In the appendix of the thesis (see below) I show that also according to Adam Smith ([1790] 1976), generally considered as the founding father of modern economic science, economic theory and practice have to be grounded on pillars of ethical nature. Within this general context, I explore how the ethical system developed by John Rawls (1999)—one of the most distinguished philosophers of the second half of the 20th century—could shape economic theory and practice.

Specifically, I apply the social contract model developed by John Rawls (1999) to three domains which have an economic dimension: European Union integration, tax compliance, and environmental sustainability. In particular, the thesis aims to answer the following overarching research question: Is an impartial and non-binding agreement, conceived within a Rawlsian framework, sufficient to generate fair and stable redistributive institutions? Through the lens of several tools provided by the Rawlsian theory (the impartiality condition guaranteed by a hypothetical agreement behind a veil of ignorance, the difference principle, the sense of justice, agreement stability, the just saving principle) this general research question is then addressed and inflected according to the specific economic domains mentioned above.

Chapter 1 of the thesis zooms in on the European Union integration process and starts with ascertaining the existence of an institutional gap between European economic integration and European social integration. Thus, the chapter aims to understand if and how Rawls’s social contract theory can fill this gap. The core idea of chapter 1 is that Rawls's A Theory of Justice can provide a new and compelling perspective to justify and endorse the institution of a mechanism of wealth redistribution at the
European level. Indeed, the European Union holds the two elements that are at the basis of Rawls’s domestic theory: (a) a scheme of mutually advantageous cooperation, and (b) a thick network of institutions which constitute a basic structure. Based on this empirical evidence, I argue that a direct redistributive mechanism, designed to mirror the difference principle, should be introduced in the European Union, so that the resources generated by the European economic cooperation are redistributed in order to maximize the expectations of the (group of) least advantaged European citizen(s). I conclude chapter 1 by labelling this mechanism as a European Fiscal Union, meant as a possible tool to pool the European surplus into a unique pot before being redistributed (Klaser 2021a).

Chapter 2 applies Rawls’s social contract theory to the tax compliance domain. Based on a laboratory experiment with three levels of income and four possible tax regimes, the chapter analyses the effects on tax compliance of a non-binding agreement reached behind a veil of ignorance. In the experiment, two main hypothesis related to Rawls’s theory are tested: (1) that, behind a veil of ignorance, participants should vote for a tax regime consistent with the difference principle, and (2) that ex-post compliance should be non-decreasing over time because of the sense of justice, understood as a strong and effective desire to repeatedly act ex-post consistently with the ex-ante chosen principles without the intervention of any external mechanism of enforcement. The results of chapter 2 show that, compared to an ‘unveiled’ agreement, the Rawlsian veil of ignorance does not in general play any significant role in the choices of the experimental subjects: between the two treatments neither the chosen tax regimes (allocation of votes) nor the tax compliance dynamics (average tax compliance over time) change. However, the veil of ignorance produces slight changes in the compliance path of low-income and high-income subjects. With the veil of ignorance, the compliance trajectory of these groups, despite being monotonically decreasing, becomes flatter, showing that an impartial procedure like the impartiality of a tax regime can have an impact on specific sub-groups (Klaser and Mittone 2022).

Chapter 3 addresses the modern climate change issue within a Rawlsian framework. It especially focuses on the two main reasons that prevent international environmental agreements from being implemented after decades of negotiations, effectively reducing greenhouse gas emissions: (1) there is not a broad consensus about the distribution of costs between nations, and (2) given the absence of transnational insti-
tutions which can monitor and sanction defectors, compliance to the targets is fragile. Chapter 3 approaches the problem by means of a laboratory experiment that squarely replicates Rawls’s intergenerational social contract, whose outcome is the just saving principle—the normative rule to regulate distribution of resources across generations. According to Rawls’s model, the individuals discussing intergenerational principles behind a veil of ignorance are only contemporaries, simulating quite well the modern climate change challenge. However, because of the properties of the veil of ignorance, the individuals do not know to which generation they belong. The empirical evidence from the laboratory shows that this veil of ignorance decision-making setting introduced into a sequential dictator game effectively induces the subjects to converge on a sustainable intergenerational path. However, the voluntary compliance to the agreement (the sense of justice) still remains an open issue, because even small incentives to defect can undermine the compliance stability, and therefore, break the whole sustainable dynamic (Klaser, Sacconi, and Faillo 2021).

The general conclusions provide a common answer to the broader research question asked at the beginning of the thesis: Is an impartial and non-binding agreement, conceived within a Rawlsian framework, sufficient to generate fair and stable redistributive institutions? Given the mixed nature of the thesis (chapter 1 is theoretical, whereas chapters 2 and 3 are more empirical) the answer is also mixed. From a theoretical perspective, the application of the Rawlsian social contract theory provides a compelling frame for examining the possibility of creating stable and fair redistributive institutions. The proposed extensions of Rawls’s theoretical frame to European justice (chapter 1), tax compliance, and climate change (hypothesis formulations in chapters 2 and 3) provide plausible and harmonious solutions to the investigated problems. Therefore, the Rawlsian framework represents an attractive tool to explore different justice issues from a theoretical point of view. This is because the social contract model developed by John Rawls is expandable and adaptable to several research fields. At the same time, it is less clear whether the tools provided by the Rawlsian framework are sufficient to effectively generate stable and fair redistributive institutions. The empirical evidence, collected through the experiments in chapters 2 and 3, shows that theory is not followed by practice because most of the tested hypothesis were disproved. This represents an obstacle to implement Rawlsian institutions
in the real world, because they might be too idealized (Sen 2009). Nevertheless, these general conclusions should not be interpreted as a message to give up doing empirical research on Rawls’s social contract theory. Instead, they represent an incitement to work more in the direction of experimental research, to understand the conditions under which Rawls’s theory can produce useful practical results, because “empirical insights are necessary if one wants to apply any theory of justice in the real world” (Gaertner and Schokkaert 2012, 7).

In the appendix I further investigate the general role of ethics in economic theory and practice. To address this issue, I look at the hierarchy between ethics and economics established by Adam Smith, generally considered as the founding father of modern economic science. I show how the ethical system developed by Smith in the *Theory of Moral Sentiments* constitutes the moral framework within which Smith conceives the prototype of the *homo oeconomicus* at the basis of the more famous *Wealth of Nations*. While some literature considered the two spheres (and books) in conflict—the so-called ‘Adam Smith problem’ (Montes 2003)—I demonstrate that according to Smith, the market mechanism can work on the sole axiom of ‘self-interest’ (*Wealth of Nations*) if, and only if, we previously assume the moral capacity of ‘sympathy’ (*Theory of Moral Sentiments*). Meticulously analysing Smith’s words, I also reinforce the interpretation according to which the ‘invisible hand’ of the *Wealth of Nations*, usually meant to regulate mere market exchanges, has to be interpreted from a broader perspective as the ‘invisible hand’ of the ‘sympathetic but impartial spectator’ that disciplines human behaviour on a more general level. I conclude the appendix claiming that this precise hierarchy—ethics first, economics after—outlined by the father of modern economic science should be extended to any economic theory and practice (Klaser 2021b).

**REFERENCES**


Klaser, Klaudijo, and Luigi Mittone. 2022. “Can the Rawlsian Veil of Ignorance Foster Tax

Klaudijo Klaser is a postdoctoral scholar at the University of Trento, where he achieved his PhD degree in Economics and Management, with a focus on behavioural and experimental economics. His studies had been enriched with fruitful research visits at the University of Oxford and Harvard University. His primary areas of research are at the intersection between moral philosophy and economics, with a particular propensity toward the application of John Rawls’s social contract theory.
Contact e-mail: <klaudijo.klaser@unitn.it>
PHD THESIS SUMMARY:

VALENTINA ERASMO
PhD in Ethics and Economics and History of Economic Thought, June 2020
Università degli Studi “G. d’Annunzio” Chieti-Pescara

Amartya Sen’s (1977b, 1982, 1985b, 1987) conception of economic agents assigned a significant role to the analysis of rational economic motivations and their consequences on individual decision-making. In particular, Sen’s conception of economic agents differed from those of his contemporaries—especially those committed to the mainstream idea of *homo economicus*. This is because he argued that individuals can still be rational even when their choices result in a decrease of individual welfare. This idea was in line with Sen’s broader criticism of mainstream economics and its commitment to *homo economicus*. However, Sen did not devote any work to outline and defend his own conception of economic agents. Rather, he provided elements of this conception across his oeuvre. Thus, we may rebuild Sen’s conception of economic agents but only through an extensive reading of his writings.

Among the existing literature on Sen’s conception of economic agents, I argue that the most interesting work is that of John B. Davis (2003, 2007, 2009, 2012). As mentioned in the epigraph, Sen (2012) admits that Davis is correct to claim that his theory of justice (Sen 2009) largely depends on his understanding of economic agents, although this point is often implicit in Sen’s own writings. Some years later, Giovanola (2013) emphasized a different element compared to Davis’ works, namely that Sen’s criticism of *homo economicus* 1 leads to a re-thinking of the anthropological and ethical foundations of economics (Giovanola 2009), including his understanding of economic agents.

Intrigued by these readings of Sen, I began by asking whether Sen does have a distinct view of economic agents. I then rebuilt the origins and development of Sen’s understanding of economic agents. On the basis of

---

1 According to Giovanola (2013), Sen’s criticism of mainstream *homo economicus* principally concerns ethical individualism and neutrality in economic behaviours.
this historical approach, I argue that Sen began to develop his earlier conception of economic agents between the 1970s and the 1990s, alongside the origins and development of his capability approach.


The first chapter shows that understanding the intellectual origins of the capability approach—which is found in Sen’s work on social choice theory—is useful for the following reason: it brings attention to important questions such as collective choices, inequality, social welfare functions, and rationality, which are crucial for a reconstruction of Sen’s earlier conception of economic agents. These intellectual origins of the capability approach are located in Sen’s debates with Kenneth Arrow, John Rawls, and John Harsanyi.

In particular, Sen differed from Arrow (1951) with respect to the idea of rationality. Arrow understood rationality as logical consistency. This is characterized by moral and ethical neutrality in economic behaviour, as in the neoclassical tradition. Sen extends this with the claim that liberalism is also important to social welfare evaluation. Specifically, he emphasized the role of ‘minimal liberty’. As Gaertner, Pattanaik, and Suzumura (1992) have pointed out, Sen (1970) introduced individual rights in terms of a ‘recognized personal sphere’ that everybody should have and nobody can interfere with. This sphere is of importance in the context of his earlier conception of economic agents because it views economic agents as independent in their individual decision-making (Chung 2019).

Although Sen appreciated Rawls’ focus on justice and fairness, which comprised Rawls’ ethical and economic rationality, Sen (1974) rejected both the Rawlsian and utilitarian decision criteria as measures to study and evaluate inequalities. This is because they “run into some fairly
straightforward difficulties, since each leaves out completely one of the
two parts of the total picture” (Sen 1974, 308). In particular, Rawls leaves
out comparisons of levels of welfare, and utilitarians leave out compari-
sions of welfare gains and losses. This need to elaborate “a more complete
theory” (Sen 1974, 308) for studying and evaluating inequalities prepared
the ground for Sen’s redefinition of economic agents in terms of responsi-
sibility.

Finally, the exchange between Sen (1976, 1977a) and Harsanyi (1977)
is characterized by their debate on the representation of social welfare
functions. Harsanyi used linear social welfare functions, which are “the
arithmetic mean of the different individuals’ von Neumann Morgenstern
(vNM) utility functions” (Harsanyi 1977, 293). In contrast, Sen promoted
non-linear social welfare functions which “refer to representations other
than those emanating from the von Neumann Morgenstern system” (Sen
1977a, 299). Sen criticized Harsanyi because his theorems are “about the
representability of social preferences and not about utilitarianism, as the
latter requires a concept of utility that is not based exclusively on prefer-
ence” (Weymark 1991, 256). Indeed, when such a notion of utility is used,
utility functions might not be vNM functions, although the expected util-
ity axioms were satisfied. Sen (1977a) thus concluded that Harsanyi’s the-
orems imply that social welfare functions are rather “nonlinear […], and
hence obviously nonutilitarian” (Weymark 1991, 257). In particular, Sen
rejected linear welfare functions because they are not able to represent
inequalities. As with Sen’s debate with Rawls, his debate with Harsanyi
also helped lay the groundwork for Sen’s redefinition of economic agents
in terms of responsibility.

The second chapter claims that, simultaneously with the development
of the capability approach (Sen 1980, 1985a), Sen makes three conceptual
extensions of mainstream economics—these are (i) from welfare to well-
being; (ii) from agent to agency; (iii) from Paretian to Liberal Efficiency.2
These conceptual extensions are relevant for Sen’s redefinition of eco-
nomic agents.

The first extension—from welfare to well-being—advances Sen’s
(1987) criticism of welfarism. In a nutshell, this criticism was based on
the idea that the informational basis of social choices is inadequate. This
is why Sen argued that economics should return to ethics, emphasizing
the role of normativity in economics, for a more complex analysis of the
reality that surrounds us. This return to ethics enabled Sen to create those

---

2 For a deepening of this schematization, see Erasmo (2020).
methodological conditions for considering responsibility in individual decision-making.

The second extension—*from agent to agency*—dealt more directly with a redefinition of the economic agent by developing two key concepts. One was the concept of agency, and the other was a dualistic conception of a person in ethical calculations. On the first, Sen (1982) addressed the relationship between rights and goals, arguing for his alternative to welfarist consequentialism and constraint-based theories. This alternative is a consequentialist approach but is devoid of its welfarist contents. Sen achieves this by placing an intrinsic value on rights. In particular, the “fulfilment and nonrealization of rights are included among the goals, incorporated in the evaluation of states of affairs, and then applied to the choice of actions through consequential links [in what] will be called a goal rights system” (Sen 1982, 15). In this way, Sen elaborated his ethical and moral approach where the economic agents became individuals who have agency. Thus, they are responsible for the consequences of their behaviours, instead of being neutral economic agents in their decision-making, as in mainstream economics. This consideration of consequences in individual economic behaviour justifies why individuals are still rational even when their choices cause a decrease of individual welfare. For instance, responsible economic agents might rationally accept a decrease of their individual welfare in order to reduce inequalities.

On the dualistic conception of a person in ethical calculation, Sen (1987) distinguished agency from well-being (thanks to the extension *from welfare to well-being*). This was done by including more motivations for action than just self-interested (or selfish) ones. In this way, Sen confirmed the reintroduction of *responsibility* in economics. So, a person might be understood, on the one hand, as an *agent* with the ability to form goals, commitments, and values which could be “exercised at the individual level, or in groups, or through democratic participation” (Alkire 2005, 219). On the other hand, we may understand a person in terms of their *well-being*, which consists in being “concerned with a person’s achievement: how ‘well’ is his or her ‘being’?” (Sen 1985a, 5). This is in contrast with the tautological formulation of well-being in mainstream *homo economicus* (Sen 1987). This is because whatever *homo economicus* chooses contributes the most to the agent’s well-being as it represents the alternative that maximizes utility—solely interpreted in terms of well-being and ignoring agency. This espouses a monistic conception of a person in ethical evaluation.
The third and final extension—*from Paretian to Liberal Efficiency*—was articulated during the 1980s and the beginning of the 1990s. This consisted in Sen (1985c, 1993) carrying out an ethical and economic analysis of the market and its relationship with individual freedoms. According to Sen (1985c), the market and its mechanisms have a moral status, which he analysed through three different hypotheses, namely: (i) the moral necessity of the market; (ii) the freedom of choice deriving from the market, and (iii) its legitimation ascribable to antecedent rights. In “Markets and Freedoms”, Sen (1993) showed how the market might limit or promote freedoms, introducing the distinction between the ‘process which leads to freedom’ and ‘opportunity-freedoms’.

In this way, efficiency is understood in terms of ‘efficiency of opportunity-freedoms’. Because of the centrality of freedom to the evaluation of states of affairs, Sen was arguing for the need for more information apart from Paretian efficiency\(^3\) in the analysis of the market and its mechanisms. In particular, through the reintroduction of *rights* in economics, the centrality of market freedoms was replaced by that of individual freedoms, completing the extension *from Paretian to Liberal Efficiency*. This element is in line with Sen’s redefinitions of welfare as well-being (*from welfare to well-being* extension) and of responsible economic agents (*from agent to agency* extension). In turn, this reintroduction of rights in economics shows that Sen’s earlier conception of economic agents also leads to an anthropocentric understanding of the market because the centrality of market freedoms was replaced by that of individual freedoms.

Despite this schematization, these extensions are not discrete categories. Indeed, the categories of *normativity*, *responsibility*, and *rights* reintroducted by Sen inform each other and jointly offer a superior theory of well-being compared to that which prevails in mainstream economics. More specifically, the reintroduction of normativity in welfare economics has contributed to Sen’s conception of responsible economic agents. The reintroduction of responsibility in economics also finds its origins in the question of individual rights. Consequently, individual rights are the category from which the extension from Paretian efficiency to opportunity-freedoms began. Together, these categories offer a superior theory of well-being because they include more ethical elements in the analysis of individual decision-making than mainstream economics does.

---

\(^3\) Paretian efficiency (or Paretian optimality) occurs when resources can not be reallocated to make an individual better off without making at least one individual worse off (or without any loss thereof).
In the third chapter, I provide a synthesis and an extension of the elements that arise in the previous two sections, especially those of rationality and agency. The aim here is to locate Sen’s earlier conception of economic agents in a broader intellectual history where, on the one hand, he was simultaneously influenced by Greek philosophical tradition and the Enlightenment; while, on the other hand, he was influencing the works on relational goods and Ricoeur’s (2005a, 2005b, 2006) later works.  

Sen explains how economic rationality enables us to know individual preferences, such as the set of human needs, while ethical rationality integrates and is in conversation with this perspective (Erasmo 2019). This is possible thanks to the extension of the horizon of individual values that direct people’s choices, by considering ‘rational’ motivations that are distinct from selfish or self-referential ones, such as sympathy and commitment (Sen 1977b, 1985b). These two concepts are set out by him as follows: the former corresponds to the case in which the concern for others directly affects one’s own welfare. If the knowledge of torture of others makes you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment (Sen 1977b, 326).

Commitment may be defined “in terms of a person choosing an act that he believes will yield a lower level of personal welfare to him than an alternative that is also available to him” (Sen 1977b, 327). This consideration of others’ goals and choices in decision-making leads to a primacy of actions over outcomes (Sen 1997) and is the opposite of what one sees in the idea of *homo economicus*. Consequently, persons are just not economic agents, but protagonists of actions who are responsible for their actions, as has been previously discussed.

All these elements can be located in a broader intellectual context where Sen was simultaneously influenced by Greek philosophical tradition and the Enlightenment, while also influencing the works on relational goods, and Ricoeur’s later work in particular. Sen was especially influenced by Aristotle’s concepts of *kalokagathia* and *phronesis*, respectively ‘beautiful and good’ and ‘practical wisdom’ (Petrochilos 2002). In particular, Aristotle’s *phronesis* has the same value as Sen’s ethical rationality in decision-making in that it shows how to attain collective well-being in

---

4 In his later works, Ricoeur principally focused on recognition, understood as a process that realizes from individuals’ recognition of their self to be recognized by others. For more about Sen’s influence on Ricoeur’s later works, see Erasmo (2022).
a way different to *homo economicus*’ ethical individualism. Hence, *phronesis* is that virtue at the basis of commitment which modifies individual goals and choices by including those of others and adopting fair and virtuous behaviours.

The influence of the Enlightenment comes from Adam Smith and Marquis de Condorcet, who were the pioneers of what Rothschild (2001) calls the ‘warm Enlightenment of economic sentiments’. Here, rationality is not indifferent, uniform, and cold. According to Smith and Condorcet, rationality is ‘warm’, including motivations in decision-making other than those selfish or self-referential ones, as in Sen’s earlier conception of economic agents. In particular, Sen’s concept of commitment appears to be influenced by Smith’s (1759) ‘sympathy’. Their respective analyses have in common the same ‘plastic’ strength of social interactions (Khalil 1990) and are a criterion of moral approval and disapproval (Raphael 1985). According to Walsh (2003), this proximity makes Sen a key figure for a ‘renaissance or second phase of classical economics’ in the contemporary age.

The emphasis Sen placed on social interactions in economic relationships was influential to authors committed to relational goods, such as Nussbaum (1986). This is seen most clearly in the influence on Paul Ricoeur’s *The Course of Recognition*, where Ricoeur attempts to bridge the gap between self-recognition and mutual recognition through Sen’s ideas about agency and social capacities. Social capacities are extremely heterogeneous, and their most fully developed form equates with “rights and capabilities”, while their “anthropological ground” is “the power to act, agency” (Ricoeur 2005b, 135). Indeed, Ricoeur is the first who recognized the anthropological value5 of Sen’s capability approach and is the most authoritative voice for supporting my argument about the strong relationship between Sen’s capability approach and his conception of economic agents.

In conclusion, this historical analysis has shown that Sen elaborated his earlier conception of economic agents alongside the origins and development of his capability approach. More specifically, the study of intellectual origins of the capability approach has uncovered the relevance of a different understanding of rationality compared to that of mainstream *homo economicus* since the 1970s. In addition, the analysis of the

---

5 In ethics and economics studies, when we talk about ‘economic anthropology’, we are referring to the understanding of economic agents. For more, see Mahieu (2016) and Manzo and Vatiero (2012).
development of the capability approach and its conceptual extensions compared to mainstream economics has highlighted the pivotal role of agency and the dualistic conception of the person in ethical calculation since the 1980s. Finally, I have provided a synthesis of these two chapters, locating Sen’s earlier conception of economic agents in a broader intellectual history where, on the one hand, he was influenced by Greek Ancient philosophy and the economic Enlightenment, and on the other hand, he was influencing the works on relational goods and Ricoeur’s later works.

Certainly, this is not a definitive analysis: my PhD thesis is but a starting point for encouraging further research and adopting different perspectives about less developed themes in Sen’s thought, such as his conception of economic agents or historical analysis of his capability approach. I look forward to developing these themes in later contributions.

REFERENCES


Valentina Erasmo is “Economy of Francesco Academy” Research Fellow and member of the HES Executive Committee. She is interested in Amartya Sen’s thought, especially his understanding of economic agents. She completed her PhD in Ethics and Economics and History of Economic Thought at Università degli Studi “G. d'Annunzio” Chieti-Pescara in June 2020. In 2016, she got her MA in Management and Socio-Economic Development at the Università degli Studi “G. d'Annunzio” Chieti-Pescara. In 2013, she completed her MA in Philosophical Sciences at the Università degli Studi “G. d'Annunzio” Chieti-Pescara. She has recently been awarded with the 2021 AHE Best Early Career Research Prize for her paper entitled “Extending Capabilities Conception of the Individual in Economics: Relationality and Responsibility”.

Contact e-mail: <valentina.erasmo@unich.it>