Universite catholique de Louvain
Faculte des Sciences Economiques, Sociales, Politiques et de Communication
Departement des Sciences Economiques
Center for Operations Research and Econometrics

Welfare and Behavior Micro-Analysis of Economies with Agents Exhibiting Non- classical Preferences

Benoit Decerf

Thèse présentée en vue de l'obtention du grade de docteur en sciences économiques et de gestion

Cotutelle entre l'Université catholique de Louvain et Universitaet Bielefeld.

Composition du Jury:

Promoteurs:	Prof. Francois Maniquet (Université catholique de Louvain)
	Prof. Frank Riedel (Universitaet Bielefeld)
Membres:	Prof. Claude d'Aspremont (Université catholique de Louvain)
	Prof. Georg Kirchsteiger (Université Libre de Bruxelles)
	Prof. Jan-Henrik Steg (Universitaet Bielefeld)
	Prof. John Weymark (Vanderbilt University)

Louvain-la-Neuve, Belgique, Décembre 2015

Contents

Acknowledgments

On the edge of completing my thesis, there are two groups of people I would like to thank. The first group is made of those people without whom this thesis would have never started in the first place. Six years ago, I was working as an engineer for a consulting company in Brussels. The decision to re-orient myself and learn economics is the result of many discussions I had with friends and family around politics and economics. PPPC debates have been an efficient catalyst in this respect.

My wife Florence has been very instrumental for my re-orientation. She encouraged me to follow this new uncertain road and trusted me all along. Given my small savings at that time, her financial support increased the feasibility of undertaking a new master at KULeuven. I am very grateful to KULeuven and in particular Gerald Willmann who accepted my application for the MASE, a one-year masters, which opened the door to pursuing a PhD.

Two Professors at KULeuven played a key role for me. First is Andre Decoster who helped me so much as my master thesis supervisor. Then Dirk Van de Gaer who taught the Welfare Economics course in fall 2010. Dirk presented to his students a model by M. Fleurbaey and F. Maniquet about fair labor income taxation, which I found fascinating. It told myself that this was exactly what I wanted to do. Unfortunately, it was already done :-(. Nevertheless, I knew from whom I wanted to learn more.

Finally, I was able to start a PhD at UCL thanks to the intervention of two individuals. First is Olivier Sterck, who made me aware of teaching assistant positions at UCL. Then comes Francois Maniquet who, God alone knows why, decided to give a chance to a motivated albeit particularly ignorant master student.

The second group I would like to thank is made up of those individuals who were on my side during the thesis period. They contributed making these last four years the best period of my life. . . so far. Also, the quality of this thesis would not be as good – or much worse, you will judge for yourself – if I had done it on my own.

Without surprise, the greatest debt I have is to Francois Maniquet. I had heard he was a great supervisor. Rumors were correct. Francois is always available, whenever you need him. Just knock on his door and he will shout “YES”. He made so many crucial suggestions for improving the quality of my research, papers and presentations. Moreover, the moments spent in his office were always very pleasant. Finally, Francois also took quite some time advising me about how to start an academic career.

Frank Riedel, my German supervisor, has been very welcoming when I moved to Bielefeld for nine months. I think I will remember for a long time the afternoon we spent in his garden, discussing our research in the sun. Thanks to him and other scholars in Bielefeld, I have had a very nice and productive time in Germany. If you plan enrolling in the EDEEM exchange program for PhDs, do not listen to people who advise going to Italy or Portugal. Go to Bielefeld!

Martin Van der Linden is the co-author of two chapters in this thesis. How lucky have I been to meet him on the way! I have progressed so much because of our collaboration: I work more efficiently, use more appropriate softwares, understand economics better and have improved my communication skills. I am aslo grateful to Martin for our many discussions via Skype between Louvain, Nashville and Bielefeld.

Many other researchers helped me and I am afraid that I will not be able to provide a complete list. I am very grateful to all the Professors who accepted being part of my thesis committee (Claude d'Aspremont, Georg Kirchsteiger, Jan-Henrik Steg and John Weymark), the group of Flemish scholars that regularly visit CORE, the Professors of my graduate courses, and the other PhD candidates of CORE. In this last group, there are a few people deserving a particular mention: Mery and Aditi who became much more than colleagues; Stephane, Simon, Dirk and Lancelot with whom I shared several beers; Abdel, Valeria, Mikel and Andrea with whom I regularly ate at the Sablon; Manu, Sinem, Veronique, Juliana, Marion, Claudia, Margherita and many others that contributed to making CORE a living place. Not only scholars but also some friends helped me, such as Chloe Sarthou, Laurent De Block and Pierre Buysschaert who corrected some language mistakes in different parts of the thesis document.

I would also thank the anonymous drivers of the cars with plate number VD 551992 (CH), CY 737 BA (F), 1 HUE 897 (B) . . . together with hundreds of other drivers who took me onboard each time I hitch-hiked from Brussels to Louvain-la-Neuve. A special mention goes to the employees of the IBA company and to Louis, my most frequent driver.

Last, but certainly not least, I would like to thank again my beloved for all the sacrifice she has made and the support she has offered me during the last four years. This adventure comes to an end but another starts for us with our little Capucine who was born on 17 March 2015. I bet the coming years will prove to be even more enjoyable than those spent at CORE.

Introduction

When confronted with an economic problem, politicians in office must select the policy that – they believe – brings the best answer to the given problem. Economists are regularly consulted for advice. In order to make up their minds, economists build models of the economic problem in order to simulate the impact of each competing policy. They recommend selecting the policy leading to the best outcome according to their models. Hence, their recommendations depend heavily on the models constructed. A good model captures the key trade-offs of the economic problem at hand but is inevitably based on strong assumptions. These assumptions are acceptable if they are sufficiently realistic or if they do not affect which policy is recommended.

Making an inappropriate assumption in a model can lead to a wrong policy recommendation for two main reasons. First, the outcome simulated for a policy can be very different from the one that would be reached in reality. That is, the *positive* evaluation of policies is falsified by the inappropriate assumption. Second, an inappropriate assumption can lead to a bad policy recommendation if it falsifies the *normative* evaluation of the outcomes attached to the competing policies. Economists evaluate the desirability of each outcome using indicators and recommend the policy whose outcome is deemed the best by the relevant indicator. Even if the positive evaluation is correct, the recommendation of economists can be wrong if the indicator used makes judgments at odds with those that politicians in office would make.

A standard modeling assumption is that economic agents are self-centered, i.e. they only care about their personal outcome. Recently, behavioral experiments in laboratory or questionnaire studies have shown that this assumption is not realistic. Agents are not self-centered but rather other-regarding, i.e. they care about the relative aspects of their personal outcomes. Dropping the self-centered assumption has both positive and normative consequences. On the positive side, other-regarding agents behave differently than their self-centered counterparts. The outcome associated to a given policy can therefore be different if agents are other-regarding agents rather than self-centered. On the normative side, the mere fact that agents are other-regarding can affect which outcome is deemed to be the best.

The main normative criterion used by economists for comparing outcomes is Pareto efficiency. This anti-paternalism criterion forces the normative evaluation to respect unanimous agreements among the concerned agents. Consider for example an economy made up of Alice and Bob with a unique good, say income. Assume policy 1 leads to an income of 20 for Alice and 10 for Bob. Is it better than policy

2 that leads to 100 for Alice and 11 for Bob?

	Alice	Bob
Policy 1	20	10
Policy 2	100	11

If agents are self-centered, the Pareto criterion concludes that policy 2 is better. On the other hand, if Bob is other-regarding, he might prefer the outcome of policy 1 over that of policy 2. Hence, there would be no agreement between the two concerned agents and the Pareto criterion is silent. Chapters 1 and 2 investigate how to compare economic outcomes when agents are affected by their relative situations.

Written in collaboration with Martin Van der Linden, **Chapter 1** investigates the normative evaluation of distributions of a multi-good endowment among other-regarding agents. More precisely, agents are assumed to have jealous separable preferences of the well-being externality type. Being jealous, an agent is never positively affected when another agent's bundle increases. This assumption on individual preferences rules out altruistic agents but not self-centered agents. The separability assumption on other-regarding preferences rules out any behavioural consequences of dropping the self-centered assumption. Finally, the preferences of an agent are of the well-being externality type if this agent evaluates the bundle received by another agent by looking at how this bundle is valued by the other agent.

We show that, under mild assumptions on the preference profile, both the First and the Second Welfare Theorems hold. This implies that, if agents are allowed to trade, Pareto efficiency is not a motive for intervention in the distribution of resources among such other-regarding agents. Nevertheless, Pareto efficiency is a criterion providing only a very partial ranking of outcomes. There exist other normative criteria that help to discriminate between efficient outcomes and hence potentially motivate an intervention. This is for example the case of equality. The focus of Chapter 1 is to investigate indicators representing the normative judgments made by a central planner interested in both Pareto efficiency and equality. The main insight we obtain is that whether or not the indicator must take into account the other-regarding part of preferences depends on the definition of equality considered. On the one hand, equality of resources allows for focusing only on the self-centered part of preferences. The standard indicator derived for economies populated with self-centered agents can therefore be directly extended for the evaluation of the economies we consider. On the other hand, a less resourcist definition of equality based on the satisfaction obtained by the agents forces the indicator to account for the other-regarding part of preferences. For this second view on equality, we propose a new indicator. For economies populated with self-centered agents, the new indicator boils down to the standard indicator.

While Chapter 1 investigates normative judgments considering the well-being of all agents in a population, Chapter 2 is concerned with poverty judgments, which focus only on the well-being of agents at the bottom of the resource distribution.

Even if poverty is a multidimensional phenomenon in practice, Chapter 2 is limited to the measurement of income poverty. The resources considered are therefore one-dimensional.

There are both an absolute and a relative aspect to income poverty. The absolute aspect captures the ability an agent has to ensure "subsistence" by satisfying her basic needs. The relative aspect accounts for "social participation". For a given purchasing power, the ability of an agent to engage in the everyday life of her society depends crucially on the standards of living in her society. If an agent is not able to afford to participate in the customary activities of her society, she should be considered as socially excluded due to her lack of resources. An agent can hence be deemed poor just because of her low relative situation.

The absolute and relative aspects of income poverty have created a long-lasting debate about poverty alleviation policies. These policies can roughly be categorized into two types: redistributive policies and growth-promoting policies. Redistributive policies aim at transferring resources from the non-poor to the poor. These policies distort incentives to make effort and hence they potentially dampen economic growth. On the contrary, growth-promoting policies potentially increase inequalities, for opposite reasons. An intuition largely shared is that redistributive policies make sense in countries in which social participation is the main issue whereas growth-promoting policies make sense in countries in which subsistence is the main issue. Beyond these easy cases, the type of policy being the most appropriate to alleviate poverty is still largely debated. Economists have a hard time when comparing two different policy mixes because of the nonexistence of a poverty measure – i.e. no indicator of poverty – balancing the absolute and relative aspects of income poverty in a transparent way. As a result, economists cannot answer the question whether a given unequally distributed growth process is poverty reducing or not.

Chapter 2 is concerned with the definition of an income poverty measure that balances the absolute and relative aspects of poverty. I show that standard poverty measures provide very counter-intuitive judgments when assessing unequally distributed growth processes. More precisely, any standard measure either ignores the social participation effects of growth or fails to give a minimal priority to subsistence over social participation. In both cases, the poverty judgments are largely at odds with intuition. As a result, the conclusions drawn from standard poverty measures should always be considered with extreme caution. I propose a new measure of income poverty that is conceptually very simple, satisfies compelling properties and is decomposable between the absolute and relative aspects of poverty. In an empirical illustration using data from the World Bank, I show that my measure provides judgments in line with intuition, contrary to standard measures. Furthermore, I show that, depending on the initial importance of absolute poverty; my measure deems an unequally distributed growth process to be poverty-reducing or poverty-increasing.

The first two chapters only aim at evaluating economic outcomes and hence ignore the question of implementation. The objective of the literature on implementation is to find which rules of the economic game lead to outcomes satisfying efficiency and fairness principles, if any. In other words, being able to identify fair and efficient outcomes is just the first step; finding which policies lead to these outcomes is the second step. A central feature of the literature on implementation is that agents respond to incentives. The behavior of agents, and hence the outcome reached, depends crucially on the incentives given by the rules of the economic game.

Predicting the outcome reached in a game often requires more assumptions than just the maximizing behavior of the players. A game is a modeling tool for strategic interactions. In a strategic interaction, a player's outcome depends not only on the strategy she plays but also on the strategies played by the other players. A player can therefore not compute the strategy maximizing her outcome without making assumptions about the strategy played by the other players. Accordingly, economists cannot predict a game's outcome without making assumptions on the kind of strategy profile – a list containing one strategy for each player – that are likely to be played. These assumptions define *solution concepts*. A solution concept rules out the strategy profiles not meeting its internal consistency and rationality requirements. Equipped with the solution concepts studied by Game Theory, economists can predict which outcomes of a game are likely to be reached. Therefore, solution concepts allow economists to investigate which rules of the economic game lead to the desired outcomes, taking into account behavioral response of the agents.

Written in collaboration with Martin Van der Linden, **Chapter 3** investigates the incentives given by assignement mechanisms in the context of the allocation of school seats. Assignement mechanisms are commonly used in districts where a decentralized allocation – leaving students and schools to decide the allocation of seats without intervention – would lead to segregation issues. In these districts, an authority centralizes the allocation of the available school seats to a set of students. The mechanisms used in this context are matching algorithms taking two inputs:

1. for each student, the preferences of her parents over the accessible schools,
2. for each school, a priority ordering of students established by the authority.

Based on the parents' preferences and the priority orderings, an algorithm returns an allocation of the seats. From the parents' point of view, each mechanism defines a different game of school choice. There are indeed strategic interactions between the parents because the seat assigned to a student depends not only on the preferences declared by her parents but also on the preferences declared by other parents.

There are three properties we would like a mechanism for allocating school seats to satisfy. First, the mechanism should always generate a Pareto efficient allocation. This efficiency property requires that there never exists another allocation that is unanimously preferred by the parents to the allocation returned. Second, the mechanism should always respect the priority ordering established by the authority. This stability property requires that no student has her priority violated in a school she deems better than the school to which she is assigned by the mechanism. Finally, the mechanism should incentivize parents to reveal truthfully their preferences. This incentive compatibility property requires that no parent has the possibility to profitably manipulate the preferences she reveals. Unfortunately, there exists no mechanism for allocating school seats satisfying all three properties. This non-existence of a mechanism satisfying a set of very compelling properties arises in many economic contexts. Trade-offs between properties must hence be made in order to go beyond this impossibility. Sometimes, the mechanism one would consider to be the best satisfies none of the properties, but is "close" to satisfying each of them. The issue in the implementation literature is that a property is a binary notion, either a mechanism satisfies it or not. If two mechanisms do not satisfy a property, we ignore which of these mechanisms is the "closest" to satisfying it.

We propose a new family of criteria allowing for comparisons of mechanisms according to the properties they do not satisfy. Each criterion in this family is defined by a different solution concept. We apply two criteria in this family in order to compare the stability of school seats allocation mechanisms. These criteria are based on the solution concept *Nash equilibrium*, for the first, and *undominated strategy profiles*, for the second. The main finding of Chapter 3 is that the comparison of mechanisms with respect to a property can be reversed depending on the solution concept considered. The relevant solution concept to consider depends of course on the information available to parents. If the game of school choice is repeated, a parent can anticipate the strategies of the other parents and it is likely that they will coordinate on an equilibrium. If the game of school choice is a one shot game, parents ignore what other parents will declare and the best they can do might be to declare an undominated strategy.

The last chapter looks slightly like an outlier. Pursuing a PhD provides many opportunities to discover new material and research areas. Nevertheless, Chapter 4 is probably less of an outlier than it seems at first sight. It is concerned with strategic behavior in contexts in which another standard assumption of economic models must be relaxed.

When an agent ignores the value of a parameter influencing her decision, economists commonly assume that the agent assigns a probability distribution over all possible values that the parameter can take. In some environments however, agents have simply no clue about which probability distribution to pick. The archetypical example is that of an Ellsberg urn. An Ellsberg urn contains a finite number of balls – say ten balls – that are either black or white, but the proportion of white balls is unknown. It is hence not possible to assign a probability to the event that a ball drawn from that urn has the color white. It might still be the case that the agent has objective but imprecise information about this proportion. For example, the agent might know there are at least 7 white balls. Nevertheless, the agent is faced with multiple beliefs – or *priors* – with respect to the probability that the ball drawn is white.

	# of balls	# of white balls	Proba. white ball drawn
Classical urn	10	7	0.7
Ellsberg urn	10	≥ 7	$\{0.7, 0.8, 0.9, 1\}$

With an Ellsberg urn, an agent is faced with *ambiguity*, i.e. she has multiple priors with respect to the correct probability distribution of the ambiguous event. Decision theory has shown that when faced with ambiguity, an agent who is averse to ambiguity bases her decision on the least favorable probability distribution in her set of priors. Her decision rule is therefore of the maxmin type.

Written in collaboration with Frank Riedel, **Chapter 4** investigates the strategic implications of decision-making under ambiguity. We consider Ellsberg games, which expand the set of strategies accessible to players. Standard games allow players to use either pure or mixed strategies. A strategy is pure if it specifies exactly which accessible action a player takes at each state of the game. A player's strategy is mixed if it specifies for each state of the game a probability distribution over all her accessible actions. Mixed strategies allow a player to hide her actions to her opponents. This feature is particularly important in competitive games in which predictable players are easy to take advantage of.

In Ellsberg games, in addition to pure and mixed strategies, players can use ambiguous randomization strategies, called *Ellsberg strategies*. An Ellsberg strategy specifies, for each state of the game, a set of probability distributions over all actions accessible to the player. Ellsberg strategies allow a player to take advantage of the ambiguity aversion of her opponents. In Ellsberg games, an Ellsberg equilibrium is a strategy profile such that all players play a best response. An Ellsberg equilibrium is hence a solution concept generalizing Nash equilibrium. Ellsberg equilibria have nevertheless less predictive power than Nash equilibria. Indeed, there exist games in which some outcomes can be sustained by an Ellsberg equilibrium but not by a Nash equilibrium.

We focus on the interpretation to give to Ellsberg strategies. One interpretation is that players let their actions depend on the result of an ambiguous randomization experiment performed with a device such as an Ellsberg urn. Another interpretation of ambiguous strategies equilibria follows from the theorem we prove, which is an extension of Harsanyi's Purification Theorem to 2×2 normal form Ellsberg games. According to our theorem, Ellsberg equilibria can be interpreted as the limit of pure and mixed equilibria in a slightly disturbed version of the original game. This version is such that the game's payoffs are known up to a small ambiguous disturbance whose value is private information to the players.

Chapter 1

Fair Social Orderings with Other-regarding Preferences

(Joint with Martin Van der Linden)

1.1 Introduction

The recent developments of behavioral economics have drawn economists' attention to the other-regarding concerns that drive economic behaviors. While supporting the use of other-regarding preferences (ORP) in positive analysis, economists have traditionally been reluctant to base normative judgments on ORP. Some authors notably feared that taking ORP into account could lead to the acknowledgment of sadistic, malicious and other "antisocial" preferences (Harsanyi, 1982). It has therefore been argued that if agents happen to have ORP, preferences should be *laundered* in order to recover self-centered preferences before performing any normative evaluation.[1] Yet, others have argued that less malevolent forms of ORP should not necessarily be laundered, and that concerns for one's relative position in society might be relevant for social evaluations (Fleurbaey, 2012).

In this paper, we defend the latter position. Although ORP must be partially laundered, some information on ORP remains pertinent for social evaluations. We think that the extent to which ORP matter to a social planner should depend on the kind of normative principles one wants to implement. We illustrate this point with two examples. Consider the classical normative principles of efficiency and equality. Efficiency is embodied in Pareto axioms that capture the idea that the social planner should respect unanimous agreements in the population. We believe that this objective is not properly met if one only takes self-centered preferences into account. Consider the following two allocations in an economy with one good and two agents.

[1] The notion of "*preference laundering*" is from Goodin (1986).

Allocations	Jane	Kumiko
1	10	10
2	11	100

Suppose that Kumiko is *self-centered*, by which we mean that she cares only about her own bundle. On the other hand, Jane suffers when she receives less than Kumiko, to the point that she prefers allocation 1 to allocation 2. Clearly, one cannot claim that allocation 2 is better than allocation 1 by virtue of Pareto efficiency. Pareto efficiency is meant to embody a respect for *unanimous* agreements, and there is no such agreement to move from allocation 1 to allocation 2.[2]

Quite differently, equality principles state that it is sometimes desirable to harm one agent in order to make another better off. Equality principles take the form of transfer axioms identifying situations in which it is desirable to move resources from one agent to another. Whether or not ORP should play a role in these axioms is more questionable. Consider the following example.

Allocations	Jane	Kumiko	Henriqua
1	12	8	100
2	10	10	100

Assume that both Kumiko and Henriqua are self-centered. Again, Jane suffers when she receives less than Henriqua. If equality is understood as *equality of resources*, allocation 2 should be socially preferred to allocation 1.[3] From a resourcist point of view, equality should not bother about social sentiments, and Jane and Kumiko are therefore treated equally in allocation 2.

Conversely, if social sentiments matter for equality, one may consider that Jane and Kumiko are *not* equal in allocation 2, because one suffers from her social environment and the other does not. According to this view, differences in social sentiments may justify that Jane receives more resources than Kumiko, and allocation 1 could be socially preferred to allocation 2.

From a normative standpoint, we believe that the resourcist notion of equality is more appealing. Yet, we recognize that this is a controversial issue. There might be applications in which compensations based on differences in social sentiments are justified. Therefore, we will study the consequences of both normative positions on the social ranking of allocations.

[2] Which allocation is fairer remains a controversial issue. In questionnaire experiments, Schokkaert (1999) reports that about 60% of his sample considered it fair to give more to only one agent, even if it generates larger inequalities. This support diminishes with the magnitude of the induced inequality. In another questionnaire experiment from Konow (2001), up to 80% of the respondents considered it unfair to unequally increase the resources allocated to two agents (the percentage is significantly lower for different framings of the question). In these experiments, however, no information on ORP was given to the respondents. As a consequence, it is difficult to infer whether people would consider that a normative principle embodying unanimity should take ORP into account.

[3] Notwithstanding the fact that redistributing resources from Henriqua to Jane and Kumiko would also be a social improvement.

Related Literature

So far, the literature on welfare economics with ORP mainly focused on the consequences of ORP for the First and Second Welfare Theorems. Most papers show that Walrasian equilibria need not be Pareto efficient when agents have ORP (among others Winter (1969), Hochman and Rodgers (1971), and Archibald and Donaldson (1976)[4]). The failure of the First Welfare Theorem has encouraged authors to study the conditions under which redistributing resources could be Pareto improving. This line of research was initiated by Hochman and Rodgers (1969) for altruistic preferences and taken up by Brennan (1973) in the case of jealous preferences. Generally, the Second Welfare Theorem does not hold either with ORP (see e.g. Dufwenberg et al. (2011), Example 1). Consequently, many papers focused on identifying restricted ORP domains on which the validity of the Second Welfare Theorem is recovered (e.g. Winter (1969), Archibald and Donaldson (1976), Dufwenberg et al. (2011)).

With its emphasis on the Welfare Theorems, the current literature is mostly concerned with Pareto efficiency. In this paper, we study the combination of efficiency and equality principles. We leave aside questions of implementability through competitive equilibria to focus on devising social ordering functions (SOFs) that generate rankings of *all* possible allocations. Since Arrow (1950), the theory of ordinal social orderings has unfortunately been plagued with impossibility results. On economic domains, a recent line of research has sought to overcome these impossibilities while maintaining ordinal non-comparability (Fleurbaey and Maniquet, 2011). This is made possible by relaxing the Arrovian Independence of Irrelevant Alternatives axiom. In the case of ORP, this approach is followed by Treibich (2014) in a model with a single good. Treibich (2014) considers a conception of equality in which ORP matter for equality. Our paper tackles the multiple goods case and addresses both the resourcist and the non-resourcist case.

The paper is organized as follows. In section 1.2, we introduce the model and define the preference domains we study. In section 1.3, we analyze the consequences for SOFs of a resourcist conception of equality in which ORP do not matter. In section 1.4, we consider notions of equality taking ORP into account. Finally, in section 1.5, we discuss some difficulties in characterizing SOFs using the axioms we introduce, and present a couple of SOFs that fail to satisfy these axioms. Except for Proposition 3, complete proofs are in the Appendix.

1.2 The model

1.2.1 Definitions and Notation

We study the problem of allocating a social endowment of private goods to a finite set of agents having heterogeneous ORP. Apart from the domain of preferences, our

[4]An exception being (Gersbach and Haller, 2001, section 5.2) who identify conditions under which Walrasian equilibria are efficient in the presence of externalities.

framework is identical to Fleurbaey and Maniquet (2006). There are ℓ perfectly divisible goods, and the social endowment is a strictly positive vector $\Omega \in \mathbb{R}^{\ell}_{++}$. We consider the case $\ell \geq 2$, although we sometimes provide intuition in a one-good economy for simplicity. The consumption set of each agent is the non-negative orthant $\mathbb{R}^{\ell}_{+}$. The set of agents is $N := \{1, \dots, n\}$, with $n \geq 2$. We write $z_i \in \mathbb{R}^{\ell}_{+}$ for the bundle of agent $i \in N$. An allocation is a vector $z_N := (z_1, \dots, z_n)$ listing the bundles of all the agents in the economy. The set of allocations is $Z := \mathbb{R}^{n\ell}_{+}$. The *social environment* of agent i is the $(n-1) \times \ell$-dimensional vector z_{-i} which lists the bundles of all agents in z_N but i. Similarly, $z_{-i,j}$ denotes the bundle of everyone but i and j in allocation z_N.

Each agent i is associated with a preference relation R_i, an ordering over the set of *allocations* Z.[5] The asymmetric and symmetric parts of R_i are denoted P_i and I_i. Agents are assumed to differ only in their preferences. They are identical in every other respect, ranging from needs to legitimate claims over the social endowment. A preference profile (sometimes *profile* for short) $R_N := (R_i)_{i \in N}$ is a list of preference relations for every individual in the population. A typical domain of admissible profiles is denoted by $\mathcal{D}$.

Following Fleurbaey and Maniquet (2006), we study the construction of social orderings over the set of allocations as a function of the profile. A social ordering function $\boldsymbol{R}$ (SOF) maps every profile in some domain $\mathcal{D}$ to a ranking of all the allocations in Z. For any SOF $\boldsymbol{R}$ and any profile $R_N \in \mathcal{D}$, $\boldsymbol{R}(R_N)$ is the social ordering of the allocations associated with R_N. Again $\boldsymbol{P}(R_N)$ and $\boldsymbol{I}(R_N)$ represent the asymmetric and symmetric part of $\boldsymbol{R}(R_N)$.

1.2.2 Preference Domains

Preferences are continuous.

Preference axiom 1 (Continuity)**.** *For all $z_N \in Z$, $i \in N$, the upper-contour set $\{z'_N \in Z \mid z'_N \; R_i \; z_N\}$ and the lower-contour set $\{z'_N \in Z \mid z_N \; R_i \; z'_N\}$ are closed.*[6]

Moreover, for a fixed z_{-i}, agent's preferences are strictly monotonic in their own consumption.

Preference axiom 2 (Strict monotonicity in own consumption)**.** *For all $z_N, z'_N \in Z$, $i \in N$, if $z'_i > z_i$ and $z_{-i} = z'_{-i}$, then $z'_N \; P_i \; z_N$.*[7]

Next, following Dufwenberg et al. (2011), people's preferences *over their own bundle* do not depend on their social environment.

Preference axiom 3 (Separability)**.** *For all (z_i, z_{-i}), (z'_i, z_{-i}), (z_i, z'_{-i}), $(z'_i, z'_{-i}) \in Z$, $i \in N$,*

$$(z'_i, z_{-i}) \; R_i \; (z_i, z_{-i}) \Leftrightarrow (z'_i, z'_{-i}) \; R_i \; (z_i, z'_{-i}).$$

[5] An ordering is a complete, reflexive and transitive binary relation.
[6] The topology on $\mathbb{R}^{\ell}_{+}$ is the usual Euclidean topology.
[7] We denote vector inequalities by the usual $>, \geq, >>$.

We consider only "negative" social sentiments, excluding the possibility of altruism. No altruism says that, other things being equal, an agent cannot be strictly better off if another agent receives more resources.

Preference axiom 4 (No altruism)**.** *For all* $z_N, z'_N \in Z$*, if* $z_j \geq z'_j$ *for some* $j \in N$ *and* $z_{-j} = z'_{-j}$*, then* $z'_N \; R_i \; z_N$ *for all* $i \neq j \in N$*.*

Notice that No altruism does not exclude the possibility that agents are self-centered, i.e. that they care only about their own bundle.

The main reason we discard altruistic preferences is because of the pathological impossibilities induced by altruism when trying to combine efficiency and equality principles (as explained at the end of section 1.3.3). Without restricting altruistic preferences, there is little hope that one would be able to construct Paretian egalitarian social ordering functions at all.

From a normative point of view, disregarding altruistic preferences might be less problematic than it seems. A good deal of agents' altruism reflects *political* preferences (Fleurbaey, 2012). Agents might have altruistic political preferences and support the adoption of equality principles, while being hurt in terms of *everyday* preferences R_i when someone else gets more resources. Altruistic *political* preferences are accounted for by requiring the social orderings to satisfy equality principles. Then, when it comes to performing social evaluations, one should be careful to exclusively rely on everyday preferences in order to avoid *double counting* (Hammond, 1987).

This being said, everyday preferences R_i are likely to exhibit some altruism too. Such altruism would be disregarded if one assumes No altruism. Notice however that if one is really made better off in terms of R_i by transferring some of her resources, she will do so by herself. Conversely, an agent will typically not be able to alter other individuals' resources in order to alleviate their negative other-regarding feelings. Thus coping with envy requires an intervention from the social planner, whereas altruism is more easily taken care of by the agents themselves.[8] This may justify focusing primarily on preferences satisfying No altruism.

We denote by $\mathcal{R}$ the domain of profiles satisfying Continuity, Strict monotonicity in own consumption, Separability, and No altruism. For most applications, $\mathcal{R}$ is too large and it is useful to consider further restrictions. A common approach in the literature consists in partly specifying the way in which agents are affected by their social environment.

By Separability, every preference relation over allocations R_i is associated with a unique preference relation over bundles R_i^{int}. We call R_i^{int} the *internal* preferences of agent i.

Definition 1 (Internal preferences)**.** *For all* $z_i, z'_i \in \mathbb{R}^{\ell}_{+}$*,*

$$z'_i \; R_i^{int} \; z_i \Leftrightarrow (z'_i, z_{-i}) \; R_i \; (z_i, z_{-i}), \; \textit{for some } z_{-i}.$$

[8]One might still worry that overlooking altruism would limit the scope for Pareto improving redistribution. In effect, such redistributions might not be self-enforced by the agents due to coordination failures (see Dufwenberg et al. (2011), Example 2, or Warr (1982)).

A traditional way to model consumption externalities is to assume that agent i's preferences depend only on her *internal* preferences over her own bundle, and on other agents' *internal* preferences over their bundles. In utility terms, i's utility will be a function of her internal utility and the internal utility of the other agents, as in the following example.[9] By Continuity, internal preference relations are continuous and can be represented by an *internal utility function* which is denoted by m_i. An example of an ORP that satisfies our assumptions is provided by Example 1.

Example 1. *adapted from (Fehr and Schmidt, 1999)*

$$U_i(z_N) = m_i(z_i) - \frac{\alpha_i}{n-1} \sum_{j \neq i} \max[m_j(z_j) - m_i(z_i), 0].$$

In preference terms, other things being equal, i should be indifferent to a change in j's bundle as long as j is internally indifferent to the change in her bundle.

Preference axiom 5 (Well-being externality). *For all $z_N, z'_N \in Z$, if $z'_i \; I_i^{int} \; z_i$ for all $i \in N$, then $z'_N \; I_i \; z_N$ for all $i \in N$.*

The preferences represented by the utility function in Example 1 satisfy Well-being externality. We denote by $\mathcal{R}^{WBE}$ the restriction of $\mathcal{R}$ to profiles that satisfy Well-being externality.

Well-being externality is a strong assumption, but it can be defended by preference laundering considerations. Think of an economy with two goods: beef and carrots. Assume that Jane is a self-centered vegetarian, whereas Kumiko likes beef and has ORP. Now consider the two following allocations.

z_N	beef	carrots
Jane	10	2
Kumiko	10	10

z'_N	beef	carrots
Jane	15	2
Kumiko	10	10

As Jane is self-centered and vegetarian, she is indifferent between z_N and z'_N. Suppose that Kumiko's preferences do not satisfy Well-being externality and she prefers z_N to z'_N because she envies the extra beef that Jane gets in z'_N. Even if these are Kumiko's actual preferences, it is debatable whether a social planner should take them into account. It may sound odd to consider a meat lover worse off because of the meat received by a vegetarian. Instead, one may think that an agent should be allowed to envy another agent's *satisfaction about their bundle*, but not that agent's actual bundle. This is precisely what Well-being externality requires (see Archibald and Donaldson (1976) for a more detailed defense of Well-being externality) .

Another way to restrict $\mathcal{R}$ is to exclude extreme forms of ORP. For instance, some $R_N \in \mathcal{R}$ are such that, at a given allocation z_N, it is impossible to find a

[9] This is the case in most models of ORP, e.g. Edgeworth (1881), Bolton and Ockenfels (2000) and Charness and Rabin (2002). See Dufwenberg et al. (2011) for more examples in a multidimensional setup.

way to distribute additional resources while making everyone better off than in z_N. Profiles that do *not* exhibit such intense forms of ORP are said to satisfy Social monotonicity (Dufwenberg et al., 2011).

Profile axiom 1 (Social monotonicity)**.** *For all $z_N \in Z$, for all $\bar{w} \in \mathbb{R}^\ell_{++}$, there exists a distribution $w \in Z$ with $\sum_i w_i = \bar{w}$ such that if $z'_N = z_N + w$, we have,*

$$z'_N \; P_i \; z_N, \text{ for all } i \in N.$$

In this paper, we introduce a different restriction on the intensity of ORP. No resource destruction unanimity requires that any allocation z'_N obtained from z_N by only destroying resources is strictly worse than z_N *for at least one agent*, no matter how the destruction of resources is split among agents.

Profile axiom 2 (No resource destruction unanimity)**.** *For all $z_N \in Z$ and all $w := (w_1, w_2, w_3, ..., w_n)$ with $w_i \in \mathbb{R}^\ell_+$ and $\sum_i w_i > 0$, there exists $j \in N$ such that*

$$(z_N + w) \; P_j \; z_N.$$

The next proposition shows the logical relation between Social monotonicity and No resource destruction unanimity on $\mathcal{R}^{WBE}$ (the proof is in Appendix 13).

Proposition 1. *On $\mathcal{R}^{WBE}$, Social monotonicity and No resource destruction unanimity are equivalent.*

We denote by $\mathcal{R}^{NRDU}$ the subdomain of profiles in $\mathcal{R}$ satisfying No resource destruction unanimity. $\mathcal{R}^{WBE-NRDU}$ is the restriction of $\mathcal{R}$ to profiles that satisfy both Well-being externality *and* No resource destruction unanimity.

Example 2. *A preference profile R_N in which the agents' preferences can be represented by utility functions of the form*

$$U_i(z_N) = m_i(z_i) - \tfrac{\alpha_i}{n-1} \textstyle\sum_{j \neq i} m_j(z_j),$$

with $0 \leq \alpha_i < 1$ for all $i \in N$ satisfies Separability, No altruism, Well-being externality and *No resource destruction unanimity, as proven in Appendix 1.7.8.*

In the next section, we study the construction of SOFs on $\mathcal{R}^{WBE-NRDU}$ when ORP matter for efficiency, but they do not matter for equality.

1.3 When ORP do not matter for equality

1.3.1 Social Ordering Axioms

In following with the argument of the introduction, we consider that ORP are relevant for efficiency axioms independently of whether ORP are taken into account by equality axioms. Therefore, we adopt the following efficiency axiom throughout.

Social ordering axiom 1 (Strong Pareto)**.** *For all* $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$,

$$\text{if } z'_N \; R_i \; z_N \text{ for all } i \in N, \text{ then } z'_N \; \boldsymbol{R}(R_N) \; z_N,$$
$$\text{if in addition there exists } j \in N \text{ such that } z'_N \; P_j \; z_N, \text{ then } z'_N \; \boldsymbol{P}(R_N) \; z_N.$$

In a multidimensional setting, one way to formalize the idea of resource equality is through the notion of bundle dominance. When j's bundle dominates k's bundle in every dimension, a Pigou-Dalton transfer from j to k should be deemed as a social improvement.

Social ordering axiom 2 (Transfer)**.** *For all* $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$, $j, k \in N$, $\Delta \in \mathbb{R}^{\ell}_{++}$, *if*

$$z_j - \Delta = z'_j >> z'_k = z_k + \Delta,$$

and $z_{-j,k} = z'_{-j,k}$, *then* $z'_N \; \boldsymbol{R}(R_N) \; z_N$.

In multidimensional frameworks with heterogeneous self-centered preferences ($R_i = R_i^{int}$ for all $i \in N$), it is well-known that *Domination among Poor* is not compatible with Pareto efficiency axioms (Fleurbaey and Trannoy, 2003).[10] This result extends readily to our domain $\mathcal{R}$, since $\mathcal{R}$ includes profiles with only self-centered preferences. To overcome this impossibility, we follow Fleurbaey and Maniquet (2011) in weakening *Domination among Poor*.[11]

A first way to do so is to restrict our notion of equality to equality *among equals*. In our framework where the only heterogeneities come from differences in preferences, this means restricting the application of *Domination among Poor* to agents having identical preferences (see Fleurbaey and Maniquet (2011) for a normative justification). As argued in the introduction, if ORP do not matter for equality, agents should not be treated differently because of differences in ORP. Thus, if two agents are identical in every respect, except for differences in the other-regarding part of their preferences, equalizing their resources should still constitute a social improvement. A social planner should then be willing to apply *Domination among Poor* between any two agents having the same *internal* preferences.

Social ordering axiom 3 (Transfer among equals INT)**.** *For all* $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$, $j, k \in N$, $\Delta \in \mathbb{R}^{\ell}_{++}$, *if*

$$z_j - \Delta = z'_j >> z'_k = z_k + \Delta,$$
$$R_j^{int} = R_k^{int},$$

and $z_{-j,k} = z'_{-j,k}$, *then* $z'_N \; \boldsymbol{R}(R_N) \; z_N$.

[10] We say that a set of axioms is *not compatible* on a domain if there exist no SOF satisfying these axioms on this domain.

[11] This means we give priority to efficiency over equality. See Sprumont (2012) for the opposite approach.

Notice that because Transfer among equals INT is independent of ORP, the desirability of a transfer does not depend on the preferences of agents who are *not involved* in the transfer. Even if every agent not involved in the transfer prefers the pre-transfer allocation z_N to the post transfer allocation z'_N because of other-regarding considerations, z'_N must still be socially preferred to z_N.

Another way to redistribute resources while maintaining compatibility with efficiency is to restrict the application of *Domination among Poor* to some region of the allocation space. Consider a transfer from j to k such that, in the post-transfer allocation z'_N, z'_j is larger than the equal-split bundle $\Omega/|N|$ in every dimension, while z'_k is smaller than $\Omega/|N|$ in every dimension. Then Equal-split transfer says that z'_N is socially at least as good as z_N.

Social ordering axiom 4 (Equal-split transfer)**.** *For all $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$, $j,k \in N$, $\Delta \in \mathbb{R}^{\ell}_{++}$, if*

$$z_j - \Delta = z'_j >> {}^{\Omega}/_{|N|} >> z'_k = z_k + \Delta,$$

and $z_{-j,k} = z'_{-j,k}$, then $z'_N \ \boldsymbol{R}(R_N) \ z_N$.

When agents do not differ in legitimate claims over the endowment, the allocation in which everyone receives an equal share ${}^{\Omega}/_{|N|}$ seems like a natural reference. Choosing ${}^{\Omega}/_{|N|}$ as a reference for transfer axioms means that ${}^{\Omega}/_{|N|}$ is viewed as an *internal* welfare lower-bound.

1.3.2 Possibility Results

In the self-centered setting, Fleurbaey and Maniquet (2011) introduced a SOF based on the egalitarian equivalence principle (Pazner and Schmeidler, 1978). This SOF consists in applying the leximin criterion to an index u_i^{Ω} measuring i's well-being as the share of the social endowment that would leave i indifferent with her current bundle. In our framework, this SOF naturally extends as follows.

Definition 2 (Internal Ω-equivalent utility)**.** *For all $z_N \in Z$, $i \in N$ and R_i satisfying Separability,*

$$u_i^{\Omega int}(z_i, R_i^{int}) = \lambda_i \Leftrightarrow z_i \ I_i^{int} \ \lambda_i \Omega.$$

The construction of the Internal Ω-equivalent utility $u_i^{\Omega int}$ is illustrated in Figure 1.1.

Let $\geq_{\ell ex}$ be the leximin operator. That is $(u'_i)_{i\in N} \geq_{\ell ex} (u_i)_{i\in N}$ if the smallest element of $(u'_i)_{i\in N}$ is greater than smallest element of $(u_i)_{i\in N}$, or they are equal and the second largest element of $(u'_i)_{i\in N}$ is greater than the second smallest element of $(u_i)_{i\in N}$, and so on.

Social ordering function 1 (Internal Ω-equivalent leximin ($\boldsymbol{R}^{\Omega \ell ex}$))**.** *For all $R_N \in \mathcal{D}$ satisfying Separability, $z_N, z'_N \in Z$*

$$z'_N \ \boldsymbol{R}^{\Omega \ell ex}(R_N) \ z_N \Leftrightarrow (u_i^{\Omega int}(z'_i, R_i^{int}))_{i\in N} \geq_{\ell ex} (u_i^{\Omega int}(z_i, R_i^{int}))_{i\in N}.$$

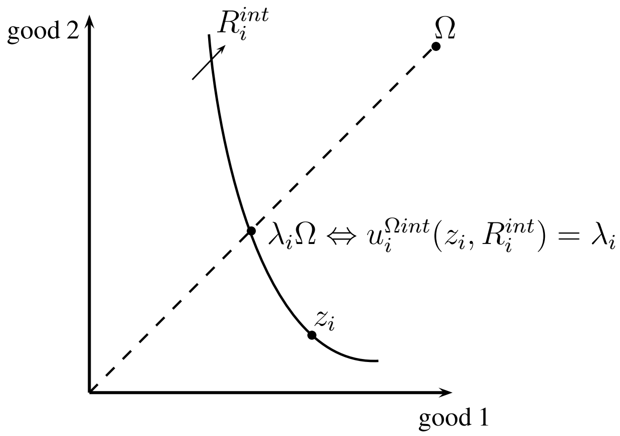

Figure 1.1: Internal Ω-equivalent utility $u_i^{\Omega int}$

A peculiar feature of $\boldsymbol{R}^{\Omega\ell ex}$ is that it can go against the preferences of the agent who is the worst-off according to $u_i^{\Omega int}$. Because $\boldsymbol{R}^{\Omega\ell ex}$ only uses information on *internal* preferences, it can be that $z'_N \; \boldsymbol{P}^{\Omega\ell ex}(R_N) \; z_N$ although z_N is preferred to z'_N by the agent with the lowest $u_i^{\Omega int}$. Nevertheless, this SOF satisfies our two equality axioms *and* Strong Pareto on $\mathcal{R}^{WBE-NRDU}$.

Proposition 2. *On the domain $\mathcal{R}^{WBE-NRDU}$, $\boldsymbol{R}^{\Omega\ell ex}$ satisfies Strong Pareto, Transfer among equals INT and Equal-split transfer.*

Proposition 2 is driven by the fact, established in the proof of Proposition 3, that on $\mathcal{R}^{WBE-NRDU}$, there cannot be a Pareto improvement unless every agent is *internally* better off. To put it in another way, the set of pairs of allocations (z_N, z'_N) such that $z'_N \; R_i \; z_N$ for all $i \in N$ is a subset of the set of pairs (z_N, z'_N) such that $z'_i \; R_i^{int} \; z_i$ for all $i \in N$. Let us define the Pareto axiom relying only on internal preferences.

Social ordering axiom 5 (Strong Pareto INT)**.** *For all $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$,*

if $z'_i \; R_i^{int} \; z_i$ for all $i \in N$, then $z'_N \; \boldsymbol{R}(R_N) \; z_N$,
if in addition there exist $j \in N$ for which $z'_j \; P_j^{int} \; z_j$ then, $z'_N \; \boldsymbol{P}(R_N) \; z_N$.

Then we have the following proposition.

Proposition 3. *On the domain $\mathcal{R}^{WBE-NRDU}$, Strong Pareto INT implies Strong Pareto.*

Proof. *We show that if an allocation z_2 Pareto dominates another allocation z_1, then allocation z_2 also dominates z_1 according to the internal preferences. This implies that an SOF that satisfies Strong Pareto INT also satisfies Strong Pareto.*

The construction of the allocations used in the proof is illustrated in Figure 1.2. Assume to the contrary that there exist z_N^1, $z_N^2 \in Z$ such that:

$$z_N^2 \; R_i \; z_N^1 \qquad \text{for all } \; i \in N, \tag{1.1}$$
$$z_j^1 \; P_j^{int} \; z_j^2 \qquad \text{for some } j \in N, \tag{1.2}$$

By No altruism, if (2) is satisfied, there exists a $k \neq j \in N$ such that $z_k^1 \; P_k^{int} \; z_k^2$. Let $S = \{i \in N | z_i^1 \; P_i^{int} \; z_i^2\}$ be the subset of agents who strictly prefer their bundles in z_N^1 to their bundle in z_N^2. In the figure, $S = \{j, k\}$ and $N \backslash S = \{g, h\}$.

Consider $z_N^3 \in Z$ constructed as follows:

$$z_i^3 >> z_i^1 \text{ and } z_i^3 \; I_i^{int} \; z_i^2 \qquad \text{for all } i \in N \backslash S,$$
$$z_i^3 = z_i^2 \qquad \text{for all } i \in S.$$

Observe that for all $i \in N$, $z_N^3 \; I_i \; z_N^2$ by Well-being externality.

Consider $z_N^4 \in Z$ constructed as follows:

$$z_i^4 = z_i^1 \qquad \text{for all } i \in N \backslash S,$$
$$z_i^4 = z_i^3 \qquad \text{for all } i \in S.$$

Observe that for all $i \in S$, $z_N^4 \; R_i \; z_N^3$ by No altruism.

Consider $z_N^5 \in Z$ constructed as follows:

$$z_i^5 >> z_i^4 \text{ and } z_i^5 \; I_i^{int} \; z_i^1 \qquad \text{for all } i \in S,$$
$$z_i^5 = z_i^1 \qquad \text{for all } i \in N \backslash S.$$

Observe that for all $i \in N$, $z_N^5 \; I_i \; z_N^1$ by Well-being externality. Given No resource destruction unanimity and the construction of z_N^4 and z_N^5, there must exist $m \in S$ such that $z_N^5 \; P_m \; z_N^4$. However, $z_N^5 \; P_m \; z_N^4$ contradicts (1.1) as we have $z_N^1 \; I_m \; z_N^5 \; P_m \; z_N^4 \; R_m \; z_N^3 \; I_m \; z_N^2$, which implies that $z_N^1 \; P_m \; z_N^2$. ■

An immediate corollary of Proposition 3 is that the set of *internally* efficient allocations is a subset of the set of efficient allocations on $\mathcal{R}^{WBE-NRDU}$ and, hence, the First Welfare Theorem holds. In their Theorem 3, Dufwenberg et al. (2011) show that the converse is true with Social monotonicity. That is, the set of efficient allocations is a subset of the set of *internally* efficient allocations and the Second Welfare Theorem holds. Because Social monotonicity is equivalent to No resource destruction unanimity on $\mathcal{R}^{WBE}$ (Proposition 1), the set of efficient allocations is identical to the set of *internally* efficient allocations on $\mathcal{R}^{WBE}$, and both the First and the Second Welfare Theorem hold.

As exemplified with $\boldsymbol{R}^{\Omega\ell ex}$, some of the SOFs satisfying Strong Pareto, Transfer among equals INT and Equal-split transfer on $\mathcal{R}^{WBE-NRDU}$ are independent of the other-regarding part of the preferences. Thus, one is *not forced* to use information on ORP in order to satisfy efficiency and resource equality. Yet, the axioms of Proposition 2 fail to uniquely characterize $\boldsymbol{R}^{\Omega\ell ex}$. One is left with some degrees of freedom in the choice of an SOF satisfying the three axioms. Then a natural

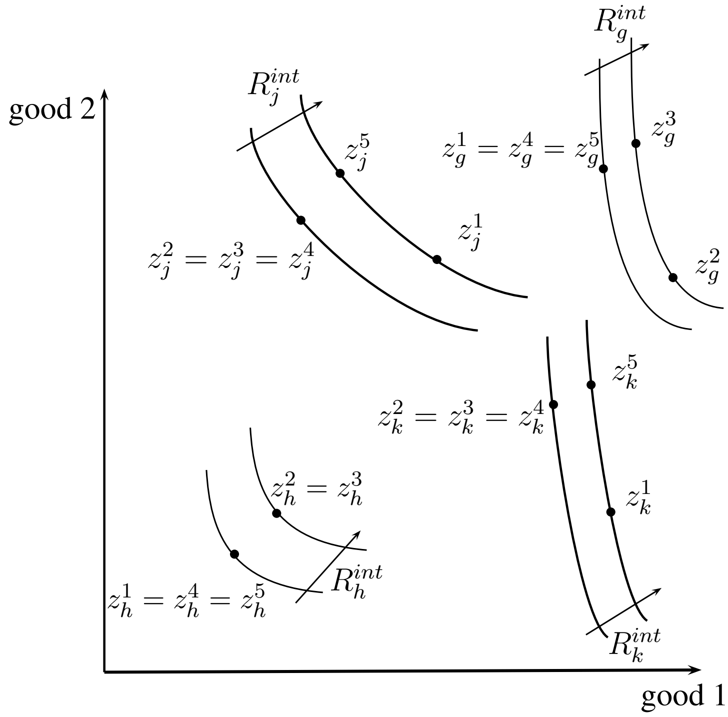

Figure 1.2: On $\mathcal{R}^{WBE-NRDU}$, there can be no Pareto improvement without an internal Pareto improvement.

question is whether this freedom allows for the use of some information on ORP. We would like to know whether $\boldsymbol{R}^{\Omega\ell ex}$ is a particular case, or whether all SOFs satisfying the axioms of Proposition 2 are necessarily independent of the other-regarding part of preferences.

An SOF is independent of the other-regarding features of preferences if it associates the same ranking of allocations to any two profiles having the same *internal* preference profile.[12]

Social ordering axiom 6 (Independence of other-regarding features)**.**
For all R_N, $R'_N \in \mathcal{D}$ and $z_N, z'_N \in Z$, if $R_i^{int} = R_i^{int\prime}$ for all $i \in N$, then

$$z'_N \; \boldsymbol{R}(R_N) \; z_N \Leftrightarrow z'_N \; \boldsymbol{R}(R'_N) \; z_N$$

The next proposition states that there exist SOFs which satisfy efficiency and equality of resources, but that are not independent of the other-regarding part of preferences.

Proposition 4. *On the domain $\mathcal{R}^{WBE-NRDU}$, there exist SOFs satisfying Strong Pareto, Transfer among equals INT and Equal-split transfer that violate Independence of other-regarding features.*

[12] Independence of other-regarding features is only used to clarify the properties of our SOFs. We do not view it as a particularly desirable axiom *a priori*.

The argument in the proof of Proposition 4 (see the Appendix) is in fact fairly general. Take any SOF $\boldsymbol{R}$ satisfying a set of axioms A and Independence of other-regarding features. If the SOF is not uniquely singled out by the set of axioms in A, there exist profiles in which the ranking of some pairs of allocations is not constrained by the axioms in A. Typically, a subset F of these unconstrained allocations are deemed indifferent by $\boldsymbol{R}$. Thus, one can construct an alternative SOF $\boldsymbol{R}^*$ which ranks allocations outside F just as $\boldsymbol{R}$, but uses some ORP information to rank allocations inside F. This new $\boldsymbol{R}^*$ will satisfy the axioms in A by construction, but violate Independence of other-regarding features. Whether any of these SOFs have desirable properties is an open question.

In the next subsection, we show that the existence of SOFs satisfying Strong Pareto, Transfer among equals INT, and Equal-split transfer is sensitive to enlargements of the domain of preferences. We also show that, once Well-being externality is dropped, it is no longer possible to construct an SOF which satisfies both Independence of other-regarding features and the other social ordering axioms.

1.3.3 Impossibility Results on Alternative Domains

First, consider the existence of SOFs satisfying Independence of other-regarding features and the other social ordering axioms. In general, if one forgoes Well-being externality, such SOFs do not exist because Strong Pareto and Independence of other-regarding features are incompatible on $\mathcal{R}^{NRDU}$ (see Proposition 5). One might wonder whether replacing Well-being externality by an alternative preference axiom permits recovering the compatibility between Strong Pareto and Independence of other-regarding features. A natural alternative to Well-being externality is Own-preference externality which states that i is indifferent to a change in j's bundle from z_j to z'_j if i herself is *internally* indifferent between z_j and z'_j.

Preference axiom 6 (Own-preference externality). *For all $z_N, z'_N \in Z$, and any $i \in N$, if $z'_j \; I_i^{int} \; z_j$ for all $j \in N$, then $z'_N \; I_i \; z_N$.*

The next example is a utility representation of a preference relation satisfying Own-preference externality.

Example 3. *from (Dufwenberg et al., 2011)*

$$U_i(z_N) = m_i(z_i) - \frac{\alpha_i}{n-1}\sum_{j\neq i} m_i(z_j).$$

Let $\mathcal{R}^{NRDU-OPE}$ be the subdomain of profiles in $\mathcal{R}^{NRDU}$ which satisfy Own-preference externality. The next proposition shows that Own-preference externality is not sufficient to recover compatibility between Strong Pareto and Independence of other-regarding features.

Proposition 5. *On the domain $\mathcal{R}^{NRDU-OPE}$, no SOF satisfies Strong Pareto and Independence of other-regarding features.*

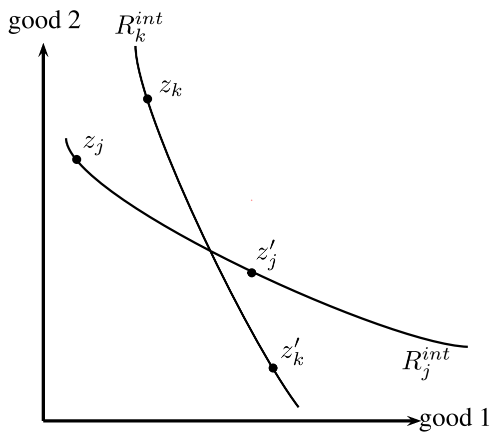

Figure 1.3: Incompatibility between Independence of other-regarding features and Strong Pareto with Own-preference externality

The proof of Proposition 5 is simple and illustrates the deep conflict between Strong Pareto and Independence of other-regarding features when preferences violate Well-being externality. Consider allocations z_N, z'_N and the internal indifference curves depicted in Figure 1.3. Let R_N be such that k is self-centered and j's preferences are as in Example 3 with $\alpha_j > 0$. Consider the symmetric profile R'_N in which j is self-centered and k's preferences are as in Example 3 with $\alpha_k > 0$. Suppose that j and k are the only two agents in the economy. Notice that in R_N, agent j prefers z'_N to z_N, because she suffers when k gets better according to her own internal preferences and $z_k \; P_j^{int} \; z'_k$. As k is indifferent between the two allocations, z'_N must be socially preferred to z_N, by Strong Pareto. The situation is symmetric in profile R'_N. In R'_N, k prefers z_N to z'_N because j is worse off in z_N according to k's internal preferences. As j is indifferent between the two allocations in R'_N, society must strictly prefer z_N to z'_N under R'_N. But Independence of other-regarding features requires that social preferences be identical in R_N and R'_N because the internal preferences are unchanged, which yields a contradiction.

As far as the compatibility of efficiency and equality axioms is concerned, forgoing Well-being externality also leads to an impossibility.

Proposition 6. *On the domain $\mathcal{R}^{NRDU-OPE}$, if $n \geq 4$, no SOF satisfies Strong Pareto and either Transfer among equals INT or Equal-split transfer.*

The impossibility in Proposition 6 is due to a common problem in combining efficiency and resource equality axioms. Problems typically arise when some Pareto improvements induce changes in bundle dominance between agents. In the self-centered case, these problems are avoided when one restricts *Domination among Poor* using either the equal preferences or the equal-split approach. This is not the case in an ORP setting unless one assumes Well-being externality, as illustrated in Figure 1.4. In some profiles in $\mathcal{R}^{NRDU-OPE}$, j may prefer z'_N to z_N although she receives strictly less of every good in z'_N. This is due to the fact that j evaluates

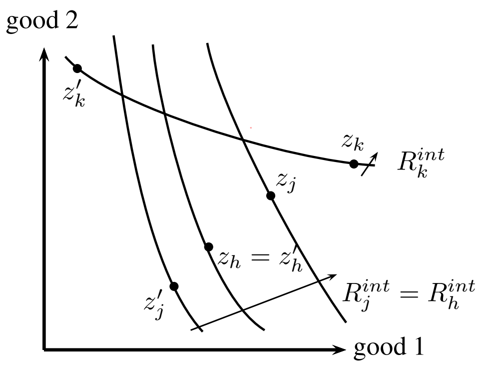

Figure 1.4: Dominance reversal between the bundles of agents j and h in $\mathcal{R}^{NRDU-OPE}$

k's situation according to her own internal preferences. Although k is internally indifferent between z'_k and z_k, j considers that k is much better off with z'_k than z_k, and this more than compensates j's own internal welfare loss. Thus all three agents are weakly better off in z'_N than in z_N, although the bundle dominance between h and j is reversed (see the proof in Appendix 1.7.4 for a precise example of a problematic preference profile).

No resource destruction unanimity is also essential for the compatibility between efficiency and resource equality, as we show in the next proposition.

Proposition 7. *On the domain $\mathcal{R}^{WBE}$, if $n \geq 4$, no SOF satisfies Strong Pareto and either Transfer among equals INT or Equal-split transfer.*

On $\mathcal{R}^{WBE}$, multiple progressive transfers can lead to an allocation which is strictly worse for *everyone* than the pre-transfer allocation. Take a one good world and the two allocations

Allocations	Jane	Kumiko	Henriqua	Madhu
z_N	1	1	4	4
z'_N	2	2	3	3

For some preference configurations violating No resource destruction unanimity, we can have $z_N \; P_i \; z'_N$ for all $i \in N$, even if z'_N is deemed at least as desirable as z_N by Equal-split transfer. Intuitively, this happens when Jane and Kumiko are harmed so much by the fact that the other receives more resources (and care so little about Henriqua and Madhu receiving less), that they both prefer allocation z_N.[13] In this case, Strong Pareto would require that $z_N \; \boldsymbol{P}(R_N) \; z'_N$ whereas equality axioms would require that $z'_N \; \boldsymbol{R}(R_N) \; z_N$, a contradiction (again see the proof in appendix 1.7.5 for a precise example of profile).

[13] In a multi-goods framework, this problem may arise even with anonymous ORP, i.e. ORP in which agents are indifferent to permutation of the other agents' bundles in the economy. See the proof in Appendix 1.7.5.

A very similar problem arises in domains that violate No altruism. On these domains, a transfer from Kumiko to Jane may be desirable according to some equality axiom, in spite of the fact that Jane prefers the pre-transfer allocation out of altruism for Kumiko. If Kumiko does not care much about Jane receiving less, there could be a unanimous agreement to go back to the pre-transfer allocation. This would make Pareto efficiency and equality axioms directly incompatible.

1.4 When ORP Matter for Equality

When ORP are taken into account in the notion of equality, the transfer axioms considered above are inappropriate in several respects. First, assume that a transfer from Jane to Kumiko harms Henriqua because of her ORP. If Henriqua is already disadvantaged as compared to Jane and Kumiko, it is not clear anymore whether such a transfer fosters equality. To avoid this kind of problem, we will further restrict our application of *Domination among Poor* to situations in which the agents not involved in the transfer are indifferent to the transfer. Such transfer axioms will be qualified as *neutral*.

Second, referring back to the example in the introduction, giving more to Jane may be justified by the fact that Jane suffers more from her social environment than Kumiko. If the difference in other-regarding feelings is "sufficiently large", then a transfer from Jane to Kumiko may not be desirable. Therefore one must define a *welfare dominance* condition determining what a "sufficiently large" difference is and adapt the transfer axioms accordingly.

Another question is how important the bundle dominance condition is to the social planner (i.e. $z_j - \Delta = z'_j >> z'_k = z_k + \Delta$ and $\Delta >> 0$). A resourcist social planner who cares about ORP might simply adapt the transfer axiom of the former section by adding a neutrality and a welfare dominance condition. We call this view a *mixed normative position.*

An alternative is for the social planner to depart completely from a resourcist notion of equality. What matters to such an observer is not whether resources are equalized, but whether, given their ORP, agents secure a similar level of well-being. This kind of social planner would only retain the neutrality and welfare dominance condition and discard any concern for bundle dominance. We call this view a *non-resourcist position.*

1.4.1 Mixed Normative Position

Let us first adapt the equal-split axiom by defining Neutral equal-split transfer. The new axiom differs from Equal-split transfer by the addition of a neutrality condition and a welfare dominance condition. The welfare dominance condition requires that the agent who benefits from the transfer prefers $(\Omega/|N|, \ldots, \Omega/|N|)$ to the post-transfer allocation, while the agent who suffers from the transfer prefers the post-transfer allocation to $(\Omega/|N|, \ldots, \Omega/|N|)$. As in Equal-split transfer, the choice of

$(\Omega/|N|, \ldots, \Omega/|N|)$ as a reference allocation reflects the idea that $(\Omega/|N|, \ldots, \Omega/|N|)$ is a natural welfare lower-bound in our distribution model.

Social ordering axiom 7 (Neutral equal-split transfer). *For all* $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$, $j, k \in N$, $\Delta \in \mathbb{R}^{\ell}_{++}$, *if*

$$z_j - \Delta = z'_j >> {}^{\Omega}/_{|N|} >> z'_k = z_k + \Delta,$$
$$z'_N \; P_j \; ({}^{\Omega}/_{|N|}, \ldots, {}^{\Omega}/_{|N|}), \; ({}^{\Omega}/_{|N|}, \ldots, {}^{\Omega}/_{|N|}) \; P_k \; z'_N,$$

$z_{-j,k} = z'_{-j,k}$, *and* $z_N \; I_i \; z'_N$ *for all* $i \neq j, k \in N$, *then* $z'_N \; \boldsymbol{R}(R_N) \; z_N$.

The next fairness axiom adapts Transfer among equals INT in a similar fashion. If ORP matter for equality, it is not always the case that one wants to redistribute between agents having the same *internal* preferences R_i^{int}. A transfer should, however, be desirable as soon as two agents have the same *global* preferences R_i. Then both agents agree that, even after the transfer, one of them is in a more favorable position.

Social ordering axiom 8 (Neutral transfer among equals). *For all* $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$, $j, k \in N$, $\Delta \in \mathbb{R}^{\ell}_{++}$, *if*

$$z_j - \Delta = z'_j >> z'_k = z_k + \Delta,$$
$$R_j = R_k,$$

$z_{-j,k} = z'_{-j,k}$ *and* $z_N \; I_i \; z'_N$ *for all* $i \neq j, k \in N$, *then* $z'_N \; \boldsymbol{R}(R_N) \; z_N$.

An immediate consequence of the way we constructed the two neutrality axioms is that they are implied by their internal counterparts. In combining resourcism with a concern for ORP, we had to make the two neutrality axioms more restrictive than Equal-split transfer and Transfer among equals INT. A corollary of Proposition 2 is therefore that, on $\mathcal{R}^{WBE-NRDU}$, $\boldsymbol{R}^{\Omega\ell ex}$ also satisfies Neutral equal-split transfer and Neutral transfer among equals. Again, $\boldsymbol{R}^{\Omega\ell ex}$ is not the only such SOF (the argument of Proposition 4 still applies). As an example, $\boldsymbol{R}^{RD\ell ex}$, which we define in the next subsection, satisfies the two neutral transfer axioms together with Strong Pareto, while violating Independence of other-regarding features.

1.4.2 Non-resourcist Position

We now analyze the consequences of departing more clearly from a resourcist conception of equality. Non-resourcist equality axioms will be called *redistributions* axioms. We first define Neutral equal-split redistribution, the non-resourcist version of Neutral equal-split transfer. With Neutral equal-split redistribution, the reduction in welfare inequalities is guaranteed by the fact that $z_N \; P_j \; z'_N \; P_j \; (\Omega/|N|, \ldots, \Omega/|N|)$, and $(\Omega/|N|, \ldots, \Omega/|N|) \; P_k \; z'_N \; P_k \; z_N$. No additional conditions are imposed.[14] In particular, bundle dominance is not required, and the resources Δ taken from k and given to j do not need to be strictly positive.

[14] Requiring that $z_N \; P_j \; z'_N$ is not essential but is meant to distinguish clearly between the Pareto efficiency and equality axioms. Without this condition, the application of Neutral equal-split redistribution would overlap with that of Strong Pareto.

Social ordering axiom 9 (Neutral equal-split redistribution). *For all* $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$, $j,k \in N$, $\Delta \in \mathbb{R}^\ell$, *if*

$$z_j - \Delta = z'_j,\ z'_k = z_k + \Delta,$$
$$z_N\ P_j\ z'_N\ P_j\ (\Omega/|N|, \dots, \Omega/|N|),\ (\Omega/|N|, \dots, \Omega/|N|)\ P_k\ z'_N\ P_k\ z_N,$$

and $z_N\ I_i\ z'_N$ *for all* $i \neq j,k \in N$, *then* $z'_N\ \boldsymbol{R}(R_N)\ z_N$.

The next fairness axiom adapts Neutral transfer among equals in a similar fashion. Let $\pi^{j,k} : Z \to Z$ denote a permutation bijection which associates every allocation z_N with the allocation $\pi^{j,k}(z_N)$ obtained by swapping j and k's bundles. In $\pi^{j,k}(z_N)$, j receives z_k, k receives z_j and the other agents get the same bundles as in z_N. In Neutral redistribution among equals, a redistribution is viewed as reducing inequalities if j and k have the same preferences, k benefits from the redistribution, and after the redistribution k is worse off than if her bundle was swapped with j's bundle.

Social ordering axiom 10 (Neutral redistribution among equals).
For all $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$, $k,j \in N$, $\Delta \in \mathbb{R}^\ell$, *if*

$$z_j - \Delta = z'_j \text{ and } z'_k = z_k + \Delta,$$
$$R_j = R_k,$$
$$z_N\ P_j\ z'_N\ P_j\ \pi^{j,k}(z'_N),\ \pi^{j,k}(z'_N)\ P_k\ z'_N\ P_k\ z_N,$$

and $z'_N\ I_i\ z_N$ *for all* $i \neq j,k \in N$, *then* $z'_N\ \boldsymbol{R}(R_N)\ z_N$.

The two redistribution axioms are satisfied by the reference distribution leximin $\boldsymbol{R}^{RD\ell ex}$. The SOF applies the leximin criterion to an index of individual well-being that we denote by u_i^{RD} for *reference distribution*. The index u_i^{RD} measures the factor λ_i which would leave i indifferent between the current allocation and $(\lambda_i \frac{\Omega}{|N|}, \frac{1}{\lambda_i}\frac{\Omega}{|N|}, \dots, \frac{1}{\lambda_i}\frac{\Omega}{|N|})$, the allocation in which i consumes $\lambda_i \frac{\Omega}{|N|}$ and everyone else consumes $\frac{1}{\lambda_i}\frac{\Omega}{|N|}$. In particular, when $\lambda_i = 1$, agent i is indifferent between the current allocation and the equal-split allocation.

Definition 3 (Reference distribution Ω-equivalent utility). *For all* $R_N \in \mathcal{D}$, *for all* $z_N \in Z$, $i \in N$,

$$u_i^{RD}(z_N, R_N) := \lambda_i \Leftrightarrow z_N\ I_i\ \left(\lambda_i \frac{\Omega}{|N|}, \frac{1}{\lambda_i}\frac{\Omega}{|N|}, \dots, \frac{1}{\lambda_i}\frac{\Omega}{|N|}\right)$$

The corresponding SOF is

Social ordering function 2 (Reference distribution leximin ($\boldsymbol{R}^{RD\ell ex}$)). *For all* $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$

$$z'_N\ \boldsymbol{R}^{RD\ell ex}(R_N)\ z_N \Leftrightarrow (u_i^{RD}(z'_N, R_N))_{i\in N} \geq_{\ell ex} (u_i^{RD}(z_N, R_N))_{i\in N}.$$

When all agents are self-centered, this SOF yields the same social ranking as $\boldsymbol{R}^{\Omega\ell ex}$.

Proposition 8. *On the domain $\mathcal{R}$, $\boldsymbol{R}^{RD\ell ex}$ satisfies Strong Pareto, Neutral redistribution among equals, and Neutral equal-split redistribution.*

Notice that we do not have to restrict the domain to $\mathcal{R}^{WBE-NRDU}$ to obtain this possibility result. This is also true in the mixed normative position. A corollary of Proposition 8 is that $\boldsymbol{R}^{RD\ell ex}$ satisfies Strong Pareto, Neutral transfer among equals, and Neutral equal-split transfer on the whole domain $\mathcal{R}$.

As is clear from the definition of $\boldsymbol{R}^{RD\ell ex}$, the SOF relies heavily on the other-regarding part of preferences. A natural question is again whether the social planner is *forced* to use such information on ORP to construct a SOF satisfying the axioms in Proposition 8. With a non-resourcist notion of equality, it turns out to be the case (contrary to the resourcist view of Section 1.3 and the mixed position of Section 1.4.1). If a social planner is willing to endorse such a strong departure from a resourcist conception of equality, she must take ORP into account in the definition of her SOF, even on $\mathcal{R}^{WBE-NRDU}$.

Proposition 9. *On the domain $\mathcal{R}^{WBE-NRDU}$, no SOF satisfies Strong Pareto, Independence of other-regarding features and either Neutral redistribution among equals or Neutral equal-split redistribution.*

The intuition of the proof is illustrated in Figure 1.5 and relies on the possibility to redistribute non-strictly positive amounts of resources between agents. The depicted economy has only two agents k and j with the same preferences. In this profile, any agent i envies the other agent only if the internal well-being of the other agent is higher than i's internal well-being.

Consider the allocations depicted in the figure. Allocation z'_N is constructed from allocation z_N by taking some of k's good 1 and giving it to j, and by taking some of j's good 2 and giving it to k. Although k is internally worse-off after the redistribution, there exist preference profiles in $\mathcal{R}^{WBE-NRDU}$ such that she prefers allocation z'_N, because her internal well-being loss is more than compensated by the internal well-being loss of j (see Appendix 1.7.7). Then by Neutral redistribution among equals and Neutral equal-split redistribution, z'_N is socially preferred to z_N. These profiles can also be constructed in such a way that z''_N is preferred by both agents to z'_N. Hence by Strong Pareto and transitivity z''_N is socially preferred to z_N. Notice however that in an alternative profile in which both agents have the same internal preferences but are self-centered, z_N is socially preferred to z''_N by efficiency. So one cannot have Independence of other-regarding features.

1.5 On characterizations and some excluded SOFs

The social ordering axioms we have used throughout the paper are not sufficient to characterize a class of SOFs. For instance, when ORP do not matter to equality, one does not have to chose the straight line going through Ω and the origin as a reference to calibrate the utility levels u_i^{Ω}. As in the self-centered case (Fleurbaey and Maniquet, 2011, Appendix 2), any other monotonic path Λ containing 0, $\frac{\Omega}{|N|}$ and

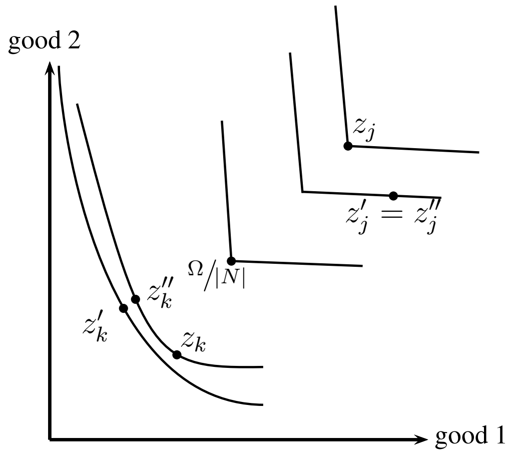

Figure 1.5: Impossibility to satisfy Independence of other-regarding features with strict non-resourcism.

Ω could be used in defining an alternative SOF $\boldsymbol{R}^{\Lambda-\ell ex}$ satisfying the above axioms. Similarly, when ORP matter to equality, one could define any two increasing and decreasing monotonic paths Λ^+ and Λ^- containing both 0 and $\frac{\Omega}{|N|}$. If these paths are use to replace the straight lines $\frac{\lambda_i \Omega}{|N|}$ and $\frac{\Omega}{\lambda_i |N|}$ in the definition of $\boldsymbol{R}^{RD\ell ex}$, the resulting SOF also satisfies the above axioms.

Characterization results using robustness axioms exist in the self-centered case (Fleurbaey and Maniquet, 2011, Chapter 5) as well as in an ORP model with a single good (Treibich, 2014). In our multidimensional setting, they are still out of reach and left for further research. We have obtained preliminary results indicating that under reasonable robustness conditions (in particular, separability conditions), transfer axioms and Strong Pareto imply that "leaky" transfers are desirable (a leaky transfer is a transfer in which what is given to the "poor" agent is less than what is taken from the "rich" agent). Such results are important building blocks for deriving characterizations along the lines of Fleurbaey and Maniquet (2011). Yet, a complete characterization remains unachieved.

This being said, our axioms do constrain significantly the realm of acceptable SOFs. In the sequel, we provide some intuition about the extent of these restrictions. We do so by introducing intuitive candidate SOFs and explaining why they are not acceptable SOFs given our axioms. We focus on the domain $\mathcal{R}^{WBE-NRDU}$ and the fairness axioms Strong Pareto and Neutral equal-split transfer.

First, we consider the family of SOFs $\boldsymbol{R}^{RD*\ell ex}$. Each member of this family is based on a particular reference distribution z^*_N which defines a specific index of individual well-being. This index measures the factor λ_i that would leave i indifferent between the current allocation and $\left(z^*_1, \ldots, z^*_{i-1}, \lambda_i z^*_i, z^*_{i+1}, \ldots, z^*_n\right)$. Notice that z^*_N need not be the equal-split allocation, nor does it need to employ the same bundle for every agent different from i.

Definition 4 (Reference distribution* z_i^*-equivalent utility). *For all $R_N \in \mathcal{D}$, for all $z_N \in Z$, $i \in N$,*

$$u_i^{RD*}(z_N, R_N) := \lambda_i \Leftrightarrow z_N \ I_i \ \left(z_1^*, \ldots, z_{i-1}^*, \lambda_i z_i^*, z_{i+1}^*, \ldots, z_n^*\right)$$

The corresponding SOF is

Social ordering function 3 (Reference distribution leximin ($\boldsymbol{R}^{RD*\ell ex}$)). *For all $R_N \in \mathcal{D}$, $z_N, z_N' \in Z$*

$$z_N' \ \boldsymbol{R}^{RD\ell ex}(R_N) \ z_N \Leftrightarrow (u_i^{RD*}(z_N', R_N))_{i\in N} \geq_{\ell ex} (u_i^{RD*}(z_N, R_N))_{i\in N}.$$

In general, the SOFs belonging to $\boldsymbol{R}^{RD*\ell ex}$ violate Neutral equal-split transfer. Indeed, a necessary condition for a leximin SOF to satisfy Neutral equal-split transfer is that for any two preference relations $R_i, R_j \in \mathcal{R}^{WBE-NRDU}$, we have $u_i^{RD*}((\Omega/|N|)_N, R_N) = u_j^{RD*}((\Omega/|N|)_N, R_N)$. Intuitively, assume that the condition is not met and that

$$j = \arg\min_{i\in N} u_i^{RD*}((\Omega/|N|)_N, R_N).$$

Then one can construct an allocation z_N close to $(\Omega/|N|)_N$ for which Neutral equal-split transfer recommends a transfer from j to another agent, although j is the worst-off according to u_i^{RD*}.

Next, we consider a special case of the former family, which we denote by $\boldsymbol{R}^{RD**\ell ex}$. This SOF takes the equal-spit allocation as the reference allocation, i.e. $z_N^* = (\Omega/|N|)_N$. As a consequence, $\boldsymbol{R}^{RD**\ell ex}$ satisfies Neutral equal-split transfer.

Definition 5 (Reference distribution** Ω-equivalent utility). *For all $R_N \in \mathcal{D}$, for all $z_N \in Z$, $i \in N$,*

$$u_i^{RD**}(z_N, R_N) := \lambda_i \Leftrightarrow z_N \ I_i \ \left(\lambda_i \frac{\Omega}{|N|}, \frac{\Omega}{|N|}, \ldots, \frac{\Omega}{|N|}\right) \tag{1.3}$$

The corresponding SOF is

Social ordering function 4 (Reference distribution** leximin ($\boldsymbol{R}^{RD**\ell ex}$)). *For all $R_N \in \mathcal{D}$, $z_N, z_N' \in Z$*

$$z_N' \ \boldsymbol{R}^{RD**\ell ex}(R_N) \ z_N \Leftrightarrow (u_i^{RD**}(z_N', R_N))_{i\in N} \geq_{\ell ex} (u_i^{RD**}(z_N, R_N))_{i\in N}.$$

$\boldsymbol{R}^{RD**\ell ex}$ is not a suitable SOF because it is not always well-specified. For certain preference profiles, there are allocations for which there exists no λ_i such that condition (1.3) is satisfied. Assume for instance that agent i is very jealous and $z_N = (z_i, 0, \ldots, 0)$. Then, there might exists no z_i' large enough such that the extra self-centered satisfaction i gets from consuming z_i' instead of z_i compensates the fact that all the other agents now receive $\frac{\Omega}{|N|}$. $\boldsymbol{R}^{RD\ell ex}$ solves this issue by allowing the reference social environment to change with λ_i (see definition 3).

Finally, we consider $\boldsymbol{R}^{RD***\ell ex}$. It is based on evaluating i's well-being index at z_N using a reference social environment in which everyone gets z_i, the bundle of i in z_N. Precisely, the index measures the factor λ_i that would leave i indifferent between the current allocation and $\left(\lambda_i \frac{\Omega}{|N|}, z_i, \ldots, z_i\right)$.

Definition 6 (Reference distribution*** Ω-equivalent utility)**.** *For all $R_N \in \mathcal{D}$, for all $z_N \in Z$, $i \in N$,*

$$u_i^{RD***}(z_N, R_N) := \lambda_i \Leftrightarrow z_N \ I_i \ \left(\lambda_i \frac{\Omega}{|N|}, z_i, \ldots, z_i\right)$$

The corresponding SOF is

Social ordering function 5 (Reference distribution*** leximin ($\boldsymbol{R}^{RD***\ell ex}$))**.** *For all $R_N \in \mathcal{D}$, $z_N, z'_N \in Z$*

$$z'_N \ \boldsymbol{R}^{RD***\ell ex}(R_N) \ z_N \Leftrightarrow (u_i^{RD***}(z'_N, R_N))_{i\in N} \geq_{\ell ex} (u_i^{RD***}(z_N, R_N))_{i\in N}.$$

The problem with $\boldsymbol{R}^{RD***\ell ex}$ is that it violates Strong Pareto. To see why, consider two allocations z_N and z'_N such that

(i) $z_j \ I_j^{int} \ z'_j$,

(ii) $z_j \neq z'_j$ and

(iii) and $z_i = z'_i$ for all $i \in N \backslash \{j\}$.

Assume that there is a unique agent $k \in N$ whose bundle affects j's ORP, and assume further that $R_j^{int} \neq R_k^{int}$. On $\mathcal{R}^{WBE-NRDU}$, we have $z_N \ I_i \ z'_N$ for all $i \in N$ by Strong Pareto. Nevertheless, $u_j^{RD***}(z_N, R_N) \neq u_j^{RD***}(z'_N, R_N)$ since j is not indifferent between an environment in which k receives z_j and another in which k receives z'_j. From these observations, it is easy to see that $\boldsymbol{R}^{RD***\ell ex}$ will not satisfy Strong Pareto.

1.6 Conclusion

When it comes to the importance of other-regarding preferences (ORP) for welfare economics, a traditional argument has been that other-regarding concerns should be laundered because taking them into account would pave the way for acknowledging antisocial and degrading traits such as malice, sadism or submissiveness (Harsanyi, 1982). In this paper, we challenge the common wisdom that preference laundering requires discarding *all* information about ORP. We argue that appropriate preference laundering should depend on the type of normative principles one is willing to implement. In particular, the same preference laundering should not necessarily apply to efficiency and equality principles. We show that it is possible to construct consistent rules for collective decisions that adopt differentiated approaches to preference launderings (Proposition 2).

Another line of research has focused on determining whether models with ORP differ significantly from self-centered models in terms of their normative properties. In general equilibrium theory, for instance, it has been repeatedly found that the First Welfare Theorem breaks down with ORP. By restricting the domain studied by Dufwenberg et al. (2011) to preferences satisfying No altruism, we obtain a different result. When one considers only "negative" ORP, efficiency in terms of

	ORP *do not* matter to equality	ORP matter to equality	
		Resourcism	Non-resourcism
$\mathcal{R}^{WBE-NRDU}$	$\boldsymbol{R}^{\Omega\ell ex}$	$\boldsymbol{R}^{\Omega\ell ex}$, $\boldsymbol{R}^{RD\ell ex}$	$\boldsymbol{R}^{RD\ell ex}$
$\mathcal{R}^{OPE-NRDU}$	$\varnothing$	$\boldsymbol{R}^{RD\ell ex}$	$\boldsymbol{R}^{RD\ell ex}$
$\mathcal{R}^{WBE}$	$\varnothing$	$\boldsymbol{R}^{RD\ell ex}$	$\boldsymbol{R}^{RD\ell ex}$

Figure 1.6: Summary of the main possibility and impossibility results. Both SOFs $\boldsymbol{R}^{\Omega\ell ex}$ and $\boldsymbol{R}^{RD\ell ex}$ use the leximin aggregator on individual indices of well-being, respectively $u_i^{\Omega int}$ and u_i^{RD}. The index $u_i^{\Omega int}$ is independent of ORP: $u_i^{\Omega int}(z_i, R_i^{int}) = \lambda \Leftrightarrow z_i \; I_i^{int} \; \lambda_i\Omega$, whereas the index u_i^{RD} makes extensive use of ORP: $u_i^{RD}(z_N, R_N) := \lambda_i \Leftrightarrow z_N \; I_i \; \left(\lambda_i \frac{\Omega}{|N|}, \frac{1}{\lambda_i}\frac{\Omega}{|N|}, \ldots, \frac{1}{\lambda_i}\frac{\Omega}{|N|}\right)$.

self-centered preferences implies efficiency for the general ORP profile (Proposition 3). As a consequence, the First Welfare Theorem still applies on this restricted ORP domain.

The literature on general equilibirum has also identified numerous ORP configurations for which the Second Welfare Theorem remains valid. This is notably the case with a condition on ORP known as Social monotonicity (Dufwenberg et al., 2011). We identify a new condition that we call No resource destruction unanimity and show that it is equivalent to Social monotonicity on a particular domain of preferences. Therefore, on this domain, efficient allocations can be implemented via competitive equilibria. However, the question of how to rank different efficient allocations remains. One way to do so is by using social-ordering functions (SOFs) based on equality principles.

If one believes that equality is a matter of resources and should not depend on ORP, we show that it is possible to rank allocations through social-ordering functions (SOFs) relying solely on self-centered preferences (Proposition 2). One of these SOFs is $\boldsymbol{R}^{\Omega\ell ex}$, a direct adaptation of the Ω-equivalent leximin SOF defined in Fleurbaey and Maniquet (2006). However, one has some freedom in the construction of SOFs satisfying equality principles. In particular, one *does not have to* exclude information on ORP from the construction of such SOFs. Whether further desirable fairness axioms would *force* the social planner to focus only on self-centered preferences is left as an open question.

Things change if the social observer departs fully from a resourcist notion of equality and considers that the harm one suffers from her social environment matters for equality. It is then impossible to construct SOFs that are independent of ORP. The reference distribution SOF $\boldsymbol{R}^{RD\ell ex}$ that we introduce provides an example of how information about ORP can be used to satisfy this second conception of equality together with efficiency.

1.7 Appendix

1.7.1 Proof of Proposition 2

On $\mathcal{R}^{WBE-NRDU}$, the social ordering function $\boldsymbol{R}^{\Omega \ell ex}$ satisfies Equal-split transfer and Strong Pareto.

Transfer among equals INT

Take any $z'_N, z_N \in Z$ s.t. $z'_N \ \boldsymbol{R}(R_N) \ z_N$ by virtue of Transfer among equals INT. By definition of Transfer among equals INT we have $z_j \gg z'_j \gg z'_k \gg z_k$, and $R_j^{int} = R_k^{int} = R^{int}$. By Strict monotonicity in own consumption, this implies

$$z_j \ P^{int} \ z'_j \ P^{int} \ z'_k \ P^{int} \ z_k,$$

which in turn means that $u_j^{\Omega int}(z_j, R_j) > u_j^{\Omega int}(z'_j, R_j) > u_k^{\Omega int}(z'_k, R_k) > u_k^{\Omega int}(z_k, R_k)$. As the $u_i^{\Omega int}$ depend only on the internal preferences, the value of $u_i^{\Omega int}$ is equal in z_N and z'_N for all $i \neq j, k$. Hence $z'_N \ \boldsymbol{R}^{\Omega \ell ex}(R_N) \ z_N$.

Equal-split transfer

The argument is identical provided that $z_j \gg z'_j \gg \frac{\Omega}{|N|} \gg z'_k \gg z_k$.

Strong Pareto

$\boldsymbol{R}^{\Omega \ell ex}$ satisfies Strong Pareto INT. By Proposition 3, on $\mathcal{R}^{WBE-NRDU}$ Strong Pareto INT implies Strong Pareto. So $\boldsymbol{R}^{\Omega \ell ex}$ satisfies Strong Pareto.

1.7.2 Proof of Proposition 4

Let Ω' be some social endowment non-proportional to Ω. Let $\boldsymbol{R}^{\Omega\Omega' \ell ex}$ be such that $z'_N \ \boldsymbol{R}^{\Omega\Omega' \ell ex}(R_N) \ z_N$ if and only if either $z'_N \ \boldsymbol{P}^{\Omega \ell ex}(R_N) \ z_N$ holds, or $z'_N \ \boldsymbol{I}^{\Omega \ell ex}(R_N) \ z_N$ and $z'_N \ \boldsymbol{R}^{\Omega' \ell ex}(R_N) \ z_N$ holds (Fleurbaey and Maniquet, 2011, chap.5).
As $\boldsymbol{R}^{\Omega \ell ex}$, this SOF satisfies all the axioms of Proposition 2 on $\mathcal{R}^{WBE-NRDU}$. Notice that $\boldsymbol{R}^{\Omega\Omega' \ell ex}$ only alters the ordering of some allocations which are deemed indifferent under $\boldsymbol{R}^{\Omega \ell ex}$.
For every profile $R_N \in \mathcal{R}^{WBE-NRDU}$, let us define the set of pairs of allocations which have different rankings under $\boldsymbol{R}^{\Omega \ell ex}$ and $\boldsymbol{R}^{\Omega\Omega' \ell ex}$, $A(R_N) = \{(z_N, z'_N) \in Z^2 \mid z_N \ \boldsymbol{I}^{\Omega \ell ex} \ z'_N$ and $z'_N \ \boldsymbol{P}^{\Omega\Omega' \ell ex} \ z_N\}$. The ranking of the pairs of allocations in $A(R_N)$ are not constrained by the set of axioms in Proposition 2. Indeed, there exist two different SOF (namely $\boldsymbol{R}^{\Omega \ell ex}$ and $\boldsymbol{R}^{\Omega\Omega' \ell ex}$) which satisfy the set of axiom in Proposition 2 and rank the pairs differently.
So consider any complete ranking $\boldsymbol{R}\big(A(R_N)\big)$ of the allocations in $A(R_N) \subset Z$ which

does not satisfy Independence of other-regarding features. For example, consider the following SOF :

$$z'_N \ \tilde{\boldsymbol{R}}^{\Omega\ell ex} \ z_N \Leftrightarrow \begin{cases} z'_N \ \boldsymbol{R}^{\Omega\ell ex}(R_N) \ z_N, & \text{if } (z_N, z'_N) \in Z \backslash A(R_N) \\ z'_N \ \boldsymbol{R}\big(A(R_N)\big) \ z_N, & \text{if } (z_N, z'_N) \in A(R_N) \end{cases}$$

By construction, $\tilde{\boldsymbol{R}}^{\Omega\ell ex}$ satisfies all the axioms in Proposition 2, but does not satisfy Independence of other-regarding features. Notice that SOF $\tilde{\boldsymbol{R}}^{\Omega\ell ex}$ is transitive by construction: it essentially follows the judgements of the transitive SOF $\boldsymbol{R}^{\Omega\ell ex}$ but ranks pairs of allocations judged indifferent by $\boldsymbol{R}^{\Omega\ell ex}$ according to the ordering $\boldsymbol{R}\big(A(R_N)\big)$ (transitive by assumption). SOF $\tilde{\boldsymbol{R}}^{\Omega\ell ex}$ hence increases the discriminative power of the SOF $\boldsymbol{R}^{\Omega\ell ex}$ without altering the strict the judgements of $\boldsymbol{R}^{\Omega\ell ex}$.

1.7.3 Proof of Proposition 5

Consider an economy with two goods z_1, z_2 and two agents j, k. The agents have the following preferences:

Preference profile R_N:

$$U_j(z_N) = m_j(z_j) - \beta m_j(z_k),$$
$$U_k(z_N) = m_k(z_k)$$

Preference profile R'_N:

$$U'_j(z_N) = m_j(z_j),$$
$$U'_k(z_N) = m_k(z_k) - \beta m_k(z_j)$$

where $R_j^{int} \neq R_k^{int}$ and $\beta > 0$. The two profiles satisfy No resource destruction unanimity as they both contain two agents, one of which is self-centered. Notice however how the profiles violates Well-being externality, as one agent's utility depends on the other's consumption through *her own* internal utility function. Consider the two allocations $z_N, z'_N \in Z$ and the internal indifference curves depicted in Figure 1.3. We have the following contradiction:

$z'_N \ P(R_N) \ z_N$ by Strong Pareto,

$z'_N \ P(R'_N) \ z_N$ by Independence of other-regarding features,

$z_N \ P(R'_N) \ z'_N$ by Strong Pareto,

1.7.4 Proof of Proposition 6

Let the economy be composed of two goods z_1, z_2 and four agents, $g, h, j, k \in N$. Agents j and k share the same internal preferences represented by the internal utility function $m = m_j = m_k$ and so do agents g and h: $m' = m_g = m_h$. Their preferences are represented by the following global utility functions:

$$\begin{aligned} U_j(z_N) &= m(z_j) - \frac{8}{11} m(z_h), \\ U_k(z_N) &= m(z_k) - \frac{8}{11} m(z_g), \\ U_g(z_N) &= m'(z_g), \\ U_h(z_N) &= m'(z_h), \end{aligned}$$

This preference profile respects both Separability and No resource destruction unanimity. Suppose the equal split bundle is $(5,5)$ and consider the following serie of allocations represented in Figure 1.7 (the level of internal utility for these allocations are as represented in the figure):

$$\begin{aligned} z_N^1 &= \Big(\underbrace{(9,9)}_{j}, \underbrace{(1,1)}_{k}, \underbrace{(1,9)}_{h}, \underbrace{(9,1)}_{g}\Big), \\ z_N^2 &= \Big(\underbrace{(8,8)}_{j}, \underbrace{(2,2)}_{k}, \underbrace{(1,9)}_{h}, \underbrace{(9,1)}_{g}\Big), \\ z_N^3 &= \Big(\underbrace{(3,3)}_{j}, \underbrace{(7,7)}_{k}, \underbrace{(9,1)}_{h}, \underbrace{(1,9)}_{g}\Big), \\ z_N^4 &= \Big(\underbrace{(4,4)}_{j}, \underbrace{(6,6)}_{k}, \underbrace{(9,1)}_{h}, \underbrace{(1,9)}_{g}\Big). \end{aligned}$$

This profile of preferences violates Well-being externality, and induces the following cycle showing that Strong Pareto and Equal-split transfer or Transfer among equals INT are not compatible.

$z_N^2 \ R(R_N) \ z_N^1$ by Equal-split transfer or by Transfer among equals INT;

$z_N^3 \ I(R_N) \ z_N^2$ by Strong Pareto;

$z_N^4 \ R(R_N) \ z_N^3$ by Equal-split transfer or by Transfer among equals INT;

$z_N^1 \ P(R_N) \ z_N^4$ by Strong Pareto, since agent k strictly prefers z_N^1.

1.7.5 Proof of Proposition 7

Let the economy be composed of two goods z_1, z_2 and four agents, g, h, j and k. Consider the following internal utility function,

$$m(z_i) = z_{i1} + z_{i2}, \tag{1.4}$$

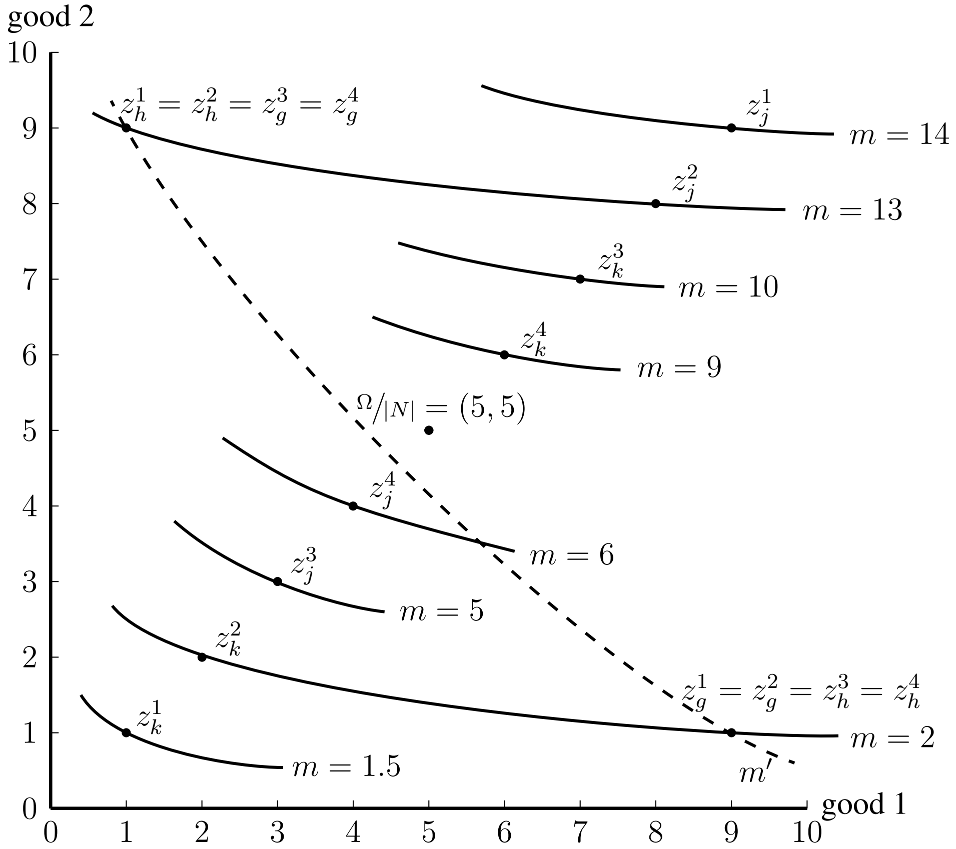

Figure 1.7: On $\mathcal{R}^{NRDU-OPE}$, no SOF satisfy Strong Pareto and Equal-split transfer.

where z_{im} is the quantity of the m-th good in i's bundle. Suppose the four agents have the following preferences.

$$U_j(z_N) = m(z_j) - \beta_j \sum_{s \neq j \in N} \frac{m(z_s)}{m(z_s)+1},$$
$$U_k(z_N) = m(z_k) - \beta_k \sum_{s \neq k \in N} \frac{m(z_s)}{m(z_s)+1},$$
$$U_g(z_N) = m(z_g),$$
$$U_h(z_N) = m(z_h),$$

where $\beta_j, \beta_k \geq 0$. This profile satisfies Well-being externality but not necessarily No resource destruction unanimity. As we show hereafter, for β_j and β_k sufficiently large, there exist allocations in which j and k would agree together to destroy part of their resources. Assume the equal split bundle is $(3,3)$ and consider the two following allocations

$$z_N = \Big(\underbrace{(1,0)}_{j}, \underbrace{(0,1)}_{k}, \underbrace{(6,5)}_{h}, \underbrace{(5,6)}_{g}\Big),$$
$$z'_N = \Big(\underbrace{(2,\epsilon)}_{j}, \underbrace{(\epsilon,2)}_{k}, \underbrace{(5,5-\epsilon)}_{h}, \underbrace{(5-\epsilon,5)}_{g}\Big),$$

for some $\epsilon > 0$ arbitrarily small.[15] Applying Transfer among equals INT or Equal-split transfer twice, we have $z'_N \ \boldsymbol{R}(R_N) \ z_N$. But we also have

$$U_j(z'_N) - U_j(z_N) \approx 1 - \beta_j \underbrace{\Big[\big(\frac{2}{3} + \frac{10}{11} + \frac{10}{11}\big) - \big(\frac{1}{2} + \frac{11}{12} + \frac{11}{12}\big)\Big]}_{=t>0},$$
$$U_k(z'_N) - U_k(z_N) \approx 1 - \beta_k \underbrace{\Big[\big(\frac{2}{3} + \frac{10}{11} + \frac{10}{11}\big) - \big(\frac{1}{2} + \frac{11}{12} + \frac{11}{12}\big)\Big]}_{=t>0},$$

where the approximation is arbitrarily accurate as ϵ tends to zero. So for $\beta_j, \beta_k > \frac{1}{t}$, $z_N \ P_j \ z'_N$ and $z_N \ P_k \ z'_N$. As $z_N \ P_g \ z'_N$ and $z_N \ P_h \ z'_N$, we have $z_N \ \boldsymbol{P}(R_N) \ z'_N$ by Strong Pareto, a contradiction.

1.7.6 Proof of Proposition 8

To prove Proposition 8, we first prove that Reference distribution Ω-equivalent utility is a utility representation of the preferences.

Lemma 1. *For any $i \in N$, u_i^{RD} is a utility representation of R_i.*

Proof. Take any $z_N, z'_N \in Z$ such that $z'_N \ R_i \ z_N$.
By definition of u_i^{RD}, $\left(\lambda'_i \frac{\Omega}{|N|}, \frac{1}{\lambda'_i}\frac{\Omega}{|N|}, \ldots, \frac{1}{\lambda'_i}\frac{\Omega}{|N|}\right) \ R_i \ \left(\lambda_i \frac{\Omega}{|N|}, \frac{1}{\lambda_i}\frac{\Omega}{|N|}, \ldots, \frac{1}{\lambda_i}\frac{\Omega}{|N|}\right)$. By Strict

[15]The ϵ is needed to meet the conditions of the transfer axiom, i.e. $\Delta >> 0$.

monotonicity in own consumption and No altruism this implies that $\lambda'_i \geq \lambda_i$, *and hence* $u_i^{RD}(z'_N, R_N) \geq u_i^{RD}(z_N, R_N)$. ■

Strong Pareto

This is a direct consequence of lemma 1 and the definition of the leximin operator.

Neutral equal-split redistribution

Let $z'_N \; \boldsymbol{R}(R_N) \; z_N$ by virtue of Neutral equal-split redistribution. By definition we have $z_N \; P_j \; z'_N$ and $z'_N \; P_k \; z_N$ so that $u_j^{RD}(z_N, R_N) > u_j^{RD}(z'_N, R_N)$ and $u_k^{RD}(z'_N) > u_k^{RD}(z_N)$ by lemma 1.
Because of the neutral character of the axiom we have that $u_i^{RD}(z_N, R_N) = u_i^{RD}(z'_N, R_N)$ for all $i \neq j, k \in N$. Also $z'_N \; P_j \; (\Omega/|N|, \ldots, \Omega/|N|)$ implies that $u_j^{RD}(z'_N, R_N) > 1$ by Strict monotonicity in own consumption and No altruism. On the other hand, $(\Omega/|N|, \ldots \Omega/|N|) \; P_k \; z'_N$ implies that $u_k^{RD}(z'_N) < 1$ for the same reason. So $u_j^{RD}(z'_N, R_N) > u_j^{RD}(z_N, R_N) > 1 > u_k^{RD}(z_N) > u_k^{RD}(z'_N)$, the desired result.

Neutral redistribution among equals

By definition of Neutral redistribution among equals we have $z_N \; P_j \; z'_N \; P_j \; \pi^{j,k}(z'_N)$, and $\pi^{j,k}(z'_N) \; P_k \; z'_N \; P_k \; z_N$. By lemma 1, and $R_j = R_k$ this implies $u_j^{RD}(z_N, R_N) > u_j^{RD}(z'_N, R_N) > u_k^{RD}(z'_N, R_N) > u_k^{RD}(z_N, R_N)$. Because of the neutral character of the axiom we have that $u_i^{RD}(z'_N, R_N) = u_i^{RD}(z_N, R_N)$ for all $i \neq j, k$.

1.7.7 Proof of Proposition 9

The proof is by contradiction. Assume there exists $\boldsymbol{R}$ satisfying Strong Pareto, Independence of other-regarding features.

Neutral equal-split redistribution

Consider a profile with two agents $j, k \in N$ having the same preferences represented by the following utility functions:

$$U_j(z_N) = \begin{cases} m(z_j) - m(z_k), & \text{if } m(z_j) < m(z_k) \\ m(z_j), & \text{if } m(z_j) \geq m(z_k) \end{cases}$$

$$U_k(z_N) = \begin{cases} m(z_k) - m(z_j), & \text{if } m(z_k) < m(z_j) \\ m(z_k), & \text{if } m(z_k) \geq m(z_j) \end{cases}$$

Notice that the induced profile satisfies No resource destruction unanimity. Let the values of the internal utility function m_i at z_N and z'_N be as represented in Figure 1.8, where $z'_k - z_k = z_j - z'_j$. We have $z'_N \; \boldsymbol{R}(R_N) \; z_N$ by Neutral equal-split redistribution. Observe that even if k is internally worse-off after the redistribution, j's internal utility loss is sufficient for k's global utility to increase. Also $z''_N \; \boldsymbol{P}(R_N) \; z'_N$ by

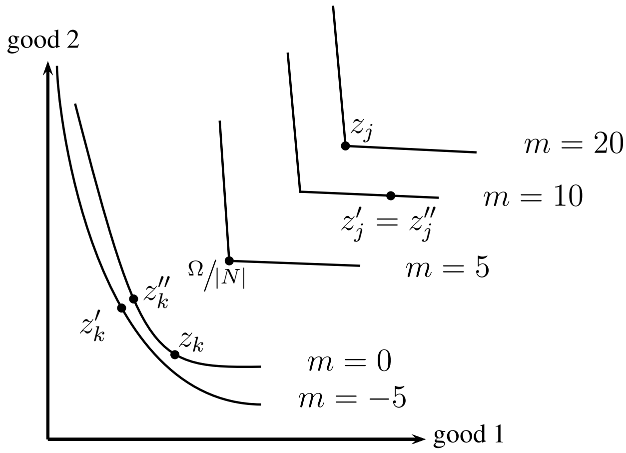

Figure 1.8: On $\mathcal{R}^{WBE-NRDU}$, no SOF satisfies Strong Pareto, Independence of other-regarding features and any of Neutral equal-split redistribution or Neutral redistribution among equals.

Strong Pareto, so $z''_N \ \boldsymbol{P}(R_N) \ z_N$ by transitivity. Now consider R'_N where $j, k \in N$ have self-centered preferences

$$u_j(z_N) = m(z_j),$$
$$u_k(z_N) = m(z_k).$$

By Strong Pareto, $z_N \ \boldsymbol{P}(R'_N) \ z''_N$, which contradicts Independence of other-regarding features.

Neutral redistribution among equals

As $R_j = R_k$, the same counterexample applies.

1.7.8 Domain of profiles given in example 2

The domain of profiles given in example 2 belongs to $\mathcal{R}^{WBE-NRDU}$. It satisfies No altruism since $\alpha_i \geq 0$, Separability because of the additively separable form of agents' ORP and Well-being externality as the utility functions is of the form $U_i(z_N) = U_i(m_i(z_i), m_j(z_j), m_k(z_k), \ldots)$. There remains to prove that No resource destruction unanimity is satisfied.

Remember that the condition means that for any $z_N \in Z$, and any $w = (w_1, w_2, \ldots, w_N)$ with $w_i \in \mathbb{R}^\ell_+$ and $\sum_i w_i = \bar{w} > 0$, we have $(z_N + w) \ P_j \ z_N$, for some $j \in N$.

Let $\Gamma_i := m_i(z_i + w_i) - m_i(z_i)$ be the internal well-being gain obtained by agent i from the distribution of w. We have that for all $i \in N$, $U_i(z_N + w) - U_i(z_N) = \Gamma_i - \frac{\alpha_i}{n-1} \sum_{j \neq i} \Gamma_j$. No resource destruction unanimity is violated if and only if we have $\Gamma_i \leq \frac{\alpha_i}{n-1} \sum_{j \neq i} \Gamma_j$ for all $i \in N$. We show that the last inequality cannot hold by contradiction.

Summing the n previous inequality yields

$$\sum_{i\in N}\Gamma_i \le \sum_{i\in N}\left[\frac{\alpha_i}{n-1}\sum_{j\neq i}\Gamma_j\right].$$

By expending the sum on the right-hand side of this inequality, on can see that

$$\sum_{i\in N}\left[\frac{\alpha_i}{n-1}\sum_{j\neq i}\Gamma_j\right] = \sum_{i\in N}\left[\Gamma_i\sum_{j\neq i}\frac{\alpha_j}{n-1}\right],$$

Let us denote the term in the parenthesis of the right-hand side $\widetilde{\Gamma}_i := \Gamma_i \sum_{j\neq i}\frac{\alpha_j}{n-1}$. Since by assumption $\alpha_i < 1$ for all $i \in N$, we have $\widetilde{\Gamma}_i < \Gamma_i$. But then the inequality cannot hold, therefore our profile must respect No resource destruction unanimity.

1.7.9 Proof of Proposition 1

No resource destruction unanimity implies Social monotonicity on $\mathcal{R}$

Take any $z_N \in Z$ and any $\bar{w} \in \mathbb{R}^\ell_{++}$. We will only need to consider the distributions of $\bar{w}$ in which each agent receives a share $\sigma_i \ge 0$ of $\bar{w}$, with $\sum_{i\in N}\sigma_i = 1$. That is the vector of additional resources for any of these distributions σ is the Kronecker product $\sigma \otimes \bar{w} = (\sigma_1\bar{w},\ldots,\sigma_n\bar{w})$, and the resulting allocation is $(z_N + \sigma\otimes\bar{w})$.

Let Σ be the $n-1$-dimensional simplex, i.e. the set of all distributions σ. Let us define the set of $\sigma \in \Sigma$ which lead to allocations that i strictly prefers to z_N,

$$B_i := \{\sigma \in \Sigma \mid (z_N + \sigma\otimes\bar{w})\ P_i\ z_N\}.$$

Similarly, let us define the indifference counter-part of B_i,

$$E_i := \{\sigma \in \Sigma \mid (z_N + \sigma\otimes\bar{w})\ I_i\ z_N\}.$$

Given the above notation, a sufficient condition for Social monotonicity to hold is $\cap_{i\in N}B_i \neq \varnothing$ and No resource destruction unanimity implies that $\sigma \in \cup_{i\in N}B_i$ for any $\sigma \in \Sigma$.

The proof is by induction. It relies on two classes of properties called Subgroup Social Monotonicity-k (SSM^k) and *Scale Invariance* (SI^{k-1}). In the induction basis, we prove SSM^2 and SI^1. The induction step then consists in proving that SSM^k and SI^{k-1} hold if SSM^{k-1} and SI^{k-2} hold. Finally, noticing that $SSM^{|N|}$ implies Social monotonicity will complete the proof.

We first introduce SSM^k, which is a version of Social monotonicity in which the new resources are only distributed to a subset of agents $N' \subseteq N$, with $|N| = k$. For any subset $N' \subseteq N$ with $|N| = k$, let $\Sigma^{N'}$ be the $k-1$-simplex, that is the distributions $\sigma \in \Sigma$ such that $\sigma_j = 0$ for all $j \in N\backslash N'$.

> **Definition 7** (Subgroup Social Monotonicity-k). *For all $N' \subseteq N$ with $|N'| = k$, for all $z_N \in Z$, and for all $\bar{w} \in \mathbb{R}^\ell_{++}$, there exists a distribution $\sigma^* \in \Sigma^{N'}$ such that,*
>
> $$(z_N + \sigma^*\otimes\bar{w})\ P_i\ z_N,\ \textit{for all } i \in N'.$$

Given the above notations, Subgroup Social Monotonicity-k holds if $\cap_{i\in N'} B_i \neq \varnothing$ for all $N' \subseteq N$ with $|N'| = k$.

Second, we introduce SI^{k-1}, which states that there is a way to distribute the complete extra resource among a subset of k agents leaving $k-1$ of them indifferent to the initial allocation, while the last agent receives a strictly positive amount of resources.

> **Definition 8** (Subgroup Indifference-(k-1))**.** *For all $N' \subseteq N$ with $|N'| = k$, for all $z_N \in Z$, for all $\bar{w} \in \mathbb{R}^{\ell}_{++}$, and for any $j \in N'$,* there exists a distribution $\sigma^* \in \Sigma^{N'}$ *with $\sigma_j > 0$ such that*
> $$(z_N + \sigma^* \otimes \bar{w})\ I_i\ z_N,\ \textit{for all } i \in N' \backslash \{j\}.$$

Given the above notations, SI^{k-1} holds if for all $N' \subseteq N$ with $|N'| = k$ and any $j \in N'$, there exists a σ with $\sigma_j > 0$ such that $\sigma \in \cap_{i \in N' \backslash \{j\}} E_i$.

<u>1. Induction basis</u>: We show that both SSM^2 and SI^1 hold under the assumptions of our domain.

First we prove SI^1. Take any two agents $j, k \in N$. For SI^1 to hold, it is enough to construct a $\sigma' \in \Sigma^{\{j,k\}}$ with $\sigma'_k > 0$ and $\sigma' \in E_j$. Notice that by Strict monotonicity in own consumption, $\Sigma^{\{j\}} \in B_j$. Also, by No altruism, $\Sigma^{\{k\}} \notin B_j$. Thus consider the continuous path that goes from $\Sigma^{\{j\}}$ to $\Sigma^{\{k\}}$ along the edge $\Sigma^{\{j,k\}}$. By Continuity, there must exists some $\sigma'' \in \Sigma^{\{j,k\}}$ with $\sigma'' \in E_j$. In order to derive a contradiction, assume $\sigma''_k = 0$. Then by No altruism and Strict monotonicity in own consumption, $\sigma'' \notin B_k$. Because $\sigma \in \Sigma^{\{j,k\}}$, $\sigma''_h = 0$ for all $h \neq j, k \in N$. Thus by No altruism and Strict monotonicity in own consumption, $\sigma'' \notin B_h$ too. But this means $\sigma'' \notin B_i$ for all $i \in N$, contradicting No resource destruction unanimity. Hence we must have $\sigma''_k > 0$, and we found the desired distribution.

Second we prove SSM^2. Take any two agents $j, k \in N$. By SI^1, there exists $\sigma' \in E_j \cap \Sigma^{\{j,k\}}$ with $\sigma'_k > 0$. As argued above, $\sigma' \notin B_h$ for all $h \neq j, k \in N$ because $\sigma' \in \Sigma^{\{j,k\}}$. But then by No resource destruction unanimity, $\sigma' \in B_k$, as otherwise no-one is strictly better at σ'. By Continuity, $B_k \cap \Sigma^{\{j,k\}}$ is open in $\Sigma^{\{j,k\}}$. Thus there exists a 1 dimensional ball $b \in \Sigma^{\{j,k\}}$ centered in σ' such that for all $\sigma \in b$ we have $\sigma \in B_k$. In particular, there exists $\sigma'' \in B_k \cap \Sigma^{\{j,k\}}$ with $\sigma''_k < \sigma'_k$, and hence $\sigma'_j < \sigma''_j$. By Strict monotonicity in own consumption, No altruism, and because the initial $\sigma' \in E_j$, we have $\sigma''\ P_j\ \sigma'$ which implies $\sigma'' \in B_j \cap B_k$, the desired result.

<u>2. Induction step</u>: If SSM^h and SI^{h-1} hold for all $h < k$, then SSM^k and SI^{k-1} hold.

First we prove SI^{k-1}. Take any $N' \subseteq N$ with $|N'| = k$ and any $j \in N'$.

Take $N' = \{1, \ldots, k\}$. We chose $\{1, \ldots, k\}$ for notational convenience and without loss of generality. Consider any distribution

$\underline{\sigma} \in \Sigma^{N' \backslash \{1\}}$ with $\underline{\sigma}_k > 0$ and $\underline{\sigma} \in \cap_{i \in \{2,\ldots,k-1\}} E_i$.

By SI^{k-2}, $\underline{\sigma}$ exists. We prove SI^{k-1} by showing the existence of a continuous path that lies in the intersection of $\Sigma^{N'}$ and $\cap_{h\in\{2,\dots,k-1\}}E_h$ and connects $\underline{\sigma}$ to some distribution in the non-empty set

$$\bar{\Sigma} := \{\bar{\sigma} \in \Sigma^{N'\setminus\{k\}} \mid \bar{\sigma}_1 > 0 \text{ and } \bar{\sigma} \in \cap_{i\in\{2,\dots,k-1\}}E_i\}.$$

If such path exists, following the same argument as in the inductions basis, it must cross E_1 at some distribution $\sigma' \in \Sigma^{N'}$, as $\underline{\sigma} \in \Sigma\setminus B_1$ whereas $\bar{\sigma} \in B_1$ and preferences are continuous. Again, if $\sigma'_k = 0$, then No resource destruction unanimity is violated as we have $\sigma' \in I_i$ for all $i \in N'\setminus\{k\}$, $\sigma'_k = 0$ and $\sigma'_j = 0$ for all $j \in N\setminus N'$. Thus σ' has the desired properties and SI^{k-1} holds.

There remains to prove the existence of such path. We construct it as the limit of a sequence of sequences $\{\gamma^n\}_{n=1}^C$, where each sequence corresponds to a different value of C. For any C, $\gamma^n \in \Sigma^{N'}$, $\gamma^1 := \underline{\sigma}$ and $\gamma^C \in \bar{\Sigma}$. These sequences are constructed in such a way that all $\gamma^n \in \cap_{i\in\{2,\dots,k-1\}}E_i$. In a nutshell, moving from γ^n to γ^{n+1} is done as follows. Let $\gamma_k^{n+1} = \gamma_k^n - \alpha\gamma_k^1$ for some $\alpha \in (0,1)$, and distribute the fraction of resource $\alpha\gamma_k^1$ among agents in $\{1,\dots,k-1\}$ in a way that leaves all agents in $\{2,\dots,k-1\}$ indifferent with γ^n. Observe that the existence of such distribution is *not* implied by SI^{k-2} as the resources are not added to the economy, but rather taken from k. If we can prove the existence of such γ^{n+1}, the properties of our domain imply that $\gamma_1^{n+1} > \gamma_1^n$. By choosing $\alpha = \frac{1}{C}$, we have $\gamma_k^C = 0$, and hence $\gamma^C \in \bar{\Sigma}$. The number C of sequences can be made arbitrarily large, which will make α arbitrarily small. Thus the sequence of sequences tends to the desired continuous path.

There remains to prove for all n that it is possible to distribute the resource $\alpha\gamma_k^1$ among agents in $\{1,\dots,k-1\}$ in such a way that $\gamma_1^{n+1} > \gamma_1^n$ and $\gamma^{n+1} \in \cap_{i\in\{2,\dots,k-1\}}E_i$. The desired distribution γ^{n+1} is constructed using the following procedure.

- Set $\gamma_k^{n+1} := \gamma_k^n - \alpha\gamma_k^1$.
- Let $\hat{\rho}^n$ be such that $\hat{\rho}_i^n := \gamma_i^n$ for all $i \in N\setminus\{k\}$ and $\hat{\rho}_k^n := \gamma_k^{n+1}$. By No altruism, we have $\hat{\rho}^n \in B_i \cup E_i$ for all $i \in N\setminus\{k\}$.
- By SSM^{k-2}, there exists a way to distribute the share of resources taken from k among agents in $\{2,\dots,k-1\}$ so as to leave them all strictly better off than in $\hat{\rho}^n$. Formally, there exists a $\rho \in \Sigma^{N'}$ with
  - $\rho_i > \gamma_i^n$ for all $i \in \{2,\dots,k-1\}$ and $\sum_{i\in\{2,\dots,k-1\}} \rho_i - \gamma_i^n = \alpha\gamma_k^1$,
  - $\rho_k = \gamma_k^{n+1}$,
  - $\rho_j = \gamma_j^n$ for all $j \in N\setminus\{2,\dots,k\}$,

  such that $\rho \in B_j$ for all $j \in \{2,\dots,k-1\}$.
- Now consider the set of distributions ρ^τ obtained by setting $\rho_j^\tau = \rho_j - \tau_j$ with $\tau_j \in (-\infty, \rho_j]$ for all $j \in \{2,\dots,k-1\}$, and transferring all the resources to agent 1, that is $\rho_1^\tau = \rho_1 + \sum_{j\in\{2,\dots,k-1\}} \tau_j$. For any $\tau := (\tau_2,\dots,\tau_{k-1})$, let ρ^τ denote the resulting allocation.

- We consider a particular subset of such distributions ρ^τ. Let

$$T := \Big\{ \tau \;\Big|\; \sum_{j\in\{2,\dots,k-1\}} \tau_j > 0, \text{ and } \rho^\tau \in B_j \cup E_j \text{ for all } j \in \{2,\dots,k-1\} \Big\}$$

  By Continuity, because $\rho \in B_j$ for all $j \in \{2,\dots,k-1\}$, there exists $\underline{\tau} >> 0$ small enough in every dimension to have $\rho^{\underline{\tau}} \in B_j$ for all $j \in \{2,\dots,k-1\}$. Thus T is non-empty.

- Now let

$$\tau^* := \arg\sup_{\tau\in T} \sum_{j\in\{2,\dots,k-1\}} \tau_j .$$

  By Continuity and the finiteness of $\{2,\dots,k-1\}$, $\cap_{j\in\{2,\dots,k-1\}}(B_j \cup E_j)$ is closed, hence $\rho^{\tau^*} \in \cap_{j\in\{2,\dots,k-1\}}(B_j \cup E_j)$. In fact, as we show below, we must have $\rho^{\tau^*} \in E_j$ for all $j \in \{2,\dots,k-1\}$. Then setting $\gamma^{n+1} = \rho^{\tau^*}$ completes the argument.

- We show that $\rho^{\tau^*} \in E_j$ for all $j \in \{2,\dots,k-1\}$ by contradiction. Assume $\rho^{\tau^*} \in B_j$ for some $j \in \{2,\dots,k-1\}$. Then by Continuity, there exists $\tilde{\tau}_j$ (small enough) such that, starting from ρ^{τ^*}, if we take $\tilde{\tau}_j$ away from j's resource and distribute it in any way among $\{1,\dots,j-1,j+1,\dots,k-1\}$, the resulting allocation remains in B_j. By SSM^{k-2} and No altruism, we can choose a redistribution $(\tilde{\tau}_1,\dots,\tilde{\tau}_{j-1},\tilde{\tau}_{j+1},\dots,\tilde{\tau}_{k-1})$ of $\tilde{\tau}_j$ that makes every agent in $\{1,\dots,j-1,j+1,\dots,k-1\}$ strictly better off at $\rho^{\tilde{\tau}}$ than at ρ^{τ^*}. Because $\rho^{\tau^*} \cap_{j\in\{2,\dots,k-1\}} (B_j \cup E_j)$, we get $\rho^{\tilde{\tau}} \in \cap_{j\in\{2,\dots,k-1\}} B_j$. Notice also that

$$\sum_{j\in\{2,\dots,k-1\}} \tilde{\tau}_j \geq \sum_{j\in\{2,\dots,k-1\}} \tau^*_j \tag{1.5}$$

  Then by the above argument, we can take away some more resources $\underline{\tilde{\tau}} >> 0$ from all agent in $\{2,\dots,k-1\}$ and redistribute them to agent 1 while remaining in $\cap_{j\in\{2,\dots,k-1\}} B_j$ (for $\underline{\tau}^* >> 0$ small enough). But this means we just found some $\hat{\tau} \in T$ such that

$$\sum_{j\in\{2,\dots,k-1\}} \hat{\tau}_j = \sum_{j\in\{2,\dots,k-1\}} \tilde{\tau}_j + \sum_{j\in\{2,\dots,k-1\}} \underline{\tilde{\tau}}_j > \sum_{j\in\{2,\dots,k-1\}} \tilde{\tau}_j , \tag{1.6}$$

  which combined with (2.19) contradicts the fact that

$$\tau^* = \arg\sup_{\tau\in T} \sum_{j\in\{2,\dots,k-1\}} \tau_j .$$

This completes the proof of SI^{k-1}.

Second we show SSM^k holds. Take any $N' \subseteq N$ with $|N'| = k$ and any $j \in N'$. By SI^{k-1}, there exists $\sigma' \in \cap_{i\in N'\setminus\{j\}} I_i \cap \Sigma^{N'}$ with $\sigma'_j > 0$. Since $B_j \cap \Sigma^{N'}$ is open in $\Sigma^{N'}$, there exists a ball $b \in \Sigma^{N'}$ centered in σ' such that for all $\sigma \in b$ we have $\sigma \in B_j$.

Take a fraction of resource from σ'_j sufficiently small such that all its distributions among agents in $N'\backslash\{j\}$ leave j in B_j. Again, by No altruism and SSM^{k-1}, there exists $\sigma'' \in b$ such that $\sigma'' \in \cap_{i\in N'} B_i$. This proves the theorem.

Social monotonicity implies No resource destruction unanimity on $\mathcal{R}^{WBE}$

The proof is by contradiction. Take any $z^{00}_N \in Z$. Assume that there exists $z^0_N \in Z$ such that

$$z^0_N \ R_i \ z^{00}_N, \qquad \text{for all } i \in N, \tag{1.7}$$
$$z^0_j < z^{00}_j, \qquad \text{for all } j \in S \subseteq N, \tag{1.8}$$
$$z^{00}_k = z^0_k, \qquad \text{for all } k \in N\backslash S, \tag{1.9}$$

so that No resource destruction unanimity is violated. We show that if Social monotonicity holds, the existence of such z^0_N implies a contradiction .

The intuition of the proof relies on the idea that by Social monotonicity, starting from z^0_N, we can start redistributing additional resources in arbitrarily small increments while making everyone better off. If we do so in the appropriate way and given the contradiction assumption (1.7), we can reach an allocation z^π_N in which for some $S_{n^*} \subset S$,

$$z^\pi_N \ P_j \ z^{00}_N, \qquad \text{for all } j \in S_{n^*}, \tag{1.10}$$
$$z^{00}_j \ I^{int}_j \ z^\pi_j, \qquad \text{for all } j \in S_{n^*}, \tag{1.11}$$
$$\text{and } z^\pi_k \ P^{int}_k \ z^{00}_k, \qquad \text{for all } k \in N\backslash S_{n^*}.$$

Then by No altruism and Well-being externality we find a series of allocations which eventually bring agents in $N\backslash S_{n^*}$ back to the bundle they had in z^{00}_N, while preserving (1.10). By Well-being externality again, we can do the same with agents in S_{n^*}, which means everyone is back at z^{00}_N. But this induces a contradiction because we then have $z^{00}_N \ P_j \ z^{00}_N$ for all $j \in S_{n^*}$.

By Social monotonicity, for any $r > 0$, there exists some $(w_1, \ldots, w_n) \in Z_+$ such that $\sum_{i\in N} w_i = (r, \ldots, r)$ and $(z^0_N + w) \ P_i \ z^0_N$ for all $i \in N$. Now consider figure 1.9. By a standard argument (see for instance (Mas-Collel et al., 1995, Proposition 3.C.1)), because preferences satisfy Strict monotonicity in own consumption and Continuity, there exists $z^{\nearrow}_N$ such that for all $i \in N$, $z^{\nearrow}_i \ I^{int}_i \ (z^0_i + w_i)$ and $z^{\nearrow}_i = z^0_i + (\gamma_i, \ldots, \gamma_i)$ for some $\gamma_i \in \mathbb{R}$. Because $w_i > 0$ by assumption and preferences satisfy Strict monotonicity in own consumption, we have $\gamma_i > 0$, for all $i \in N$.

It will be convenient to use the γ_i constructed above to define an internal utility function for agent i. For all $z_N \in Z$ and for all $i \in N$, let

$$m_i(z_i) := \gamma_i.$$

This internal utility function is defined for the particular reference allocation z^0.

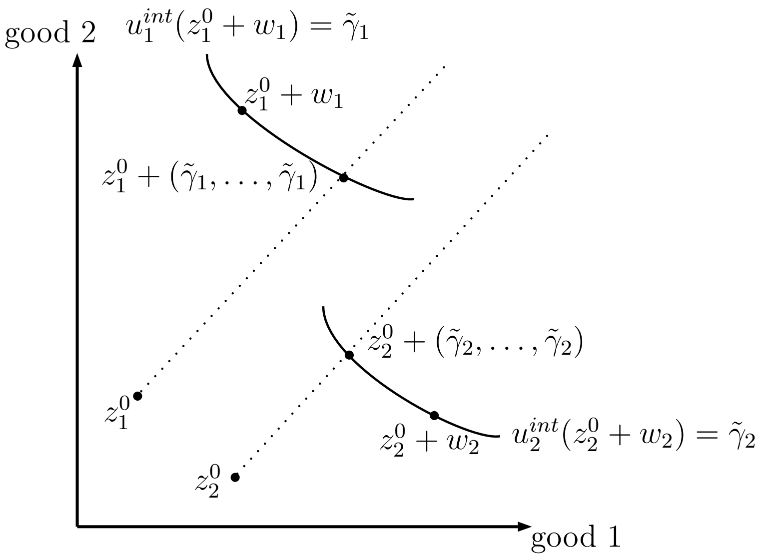

Figure 1.9: An internal utility function starting from z^0.

Notice that because preferences satisfy Separability, m_i is indeed a utility representation of i's internal preferences . Also m_i is continuous (again, see (Mas-Collel et al., 1995, Proposition 3.C.1)).

The next lemma shows that starting from any allocation, if we distribute $(r, \ldots, r)$ additional resources following Social monotonicity, there is an upper-bound to the amount of extra internal utility that can be obtained, *and that this upper-bound is strictly decreasing in r.*

To state the lemma, we need some additional notation. For any $z_N \in Z$, let $m_N(z_N) := (m_1(z_1), \ldots, m_n(z_n))$. Also, for any $r > 0$, let $M(r, z_N)$ be the set of internal utility vectors $m_N := (m_1, \ldots, m_n)$ which can be obtained from z_N by distributing $(r, \ldots, r)$ and making everyone strictly better off than at z_N. Formally

$$\begin{aligned} M(r, z_N) := \Big\{ & m_N \in \mathbb{R}^n \mid m_N = m_N(\hat{z}_N) \text{ for some } \hat{z}_N \in Z \text{ such that} \\ & \hat{z}_N = (z_N + w), \text{ for some } w \in Z_+ \text{ with } \sum_{i \in N} w_i = (r, \ldots, r), \\ & \text{and for all } i \in N,\ \hat{z}_N \ P_i \ z_N \Big\}. \end{aligned} \tag{1.12}$$

By Social monotonicity, for any $r > 0$ and any $z_N \in Z$, the set M is non-empty.

Lemma 2 (Upper-bound to extra utility strictly decreasing in r)**.**
For any $z_N \in Z$, any $r > 0$, and any $m_N \in M(r, z_N)$, $m_N \leq (m_N(z_N) + (r, \ldots, r))$.

Proof. By definition of $M(r, z_N)$ for all $m_N \in M(r, z_N)$, there exists some $\hat{z}_N \in Z$ such that $m_N = m_N(\hat{z}_N)$ and

$$\hat{z}_i \leq z_i + (r, \ldots, r), \text{ for all } i \in N \tag{1.13}$$

Now assume there exists $\tilde{m}_N \in M(r, z_N)$ *such that for some* $j \in N$, $\tilde{m}_j > (m_j(z_j) + r)$. *This implies that* $\hat{z}_j\ P_j^{int}\ \big(z_j + (r, \ldots, r)\big)$. *But by Strict monotonicity in own consumption, this means that there is some good* h *such that* $\hat{z}_{jh} > z_{jh} + r$, *contradicting* (1.13). ■

The next lemma shows that for every allocation z_N in which every agent $i \in S$ has strictly lower internal utility than at z_N^{00}, there exists another allocation z'_N which everyone prefers to z_N, in which all agents have strictly higher *internal* utility than at z_N, but in which agents $i \in S$ still have strictly lower *internal* utility than at z_N^{00}.

From now on, we will denote $z_S := (z_i)_{i \in S}$ the allocation z_N restricted to the subset of agents in S. Similarly $m_S(z_S) := (m_i(z_i))_{i \in S}$ and $m_S := (m_i)_{i \in S}$. Also let $m_N^{00} := m_N(z_N^{00})$.

Lemma 3 (Internal utility increasing but still lower than m_N^{00})**.**
Take any $z_N \in Z$ *such that* $m_S(z_S) << m_S^{00}$. *There exists* $z'_N \in Z$ *such that*

$$z'_N\ P_i\ z_N, \qquad \textit{for all } i \in N, \tag{1.14}$$

$$m_i(z_i) < m_i(z'_i), \qquad \textit{for all } i \in N, \tag{1.15}$$

$$\textit{and } m_j(z'_j) < m_j^{00}, \qquad \textit{for all } j \in S, \tag{1.16}$$

Proof. *Let*

$$r^* := \frac{\min_{s \in S} \big(m_s^{00} - m_s(z_s)\big)}{2},$$

and let $m'_n := m_N(z'_N)$ *be any element of* $M(r^*, z_N)$. *This construction is illustrated in Figure 1.10 for the case* $z_N = z_N^0$ *with two agents. By assumption* $m_S(z_S) << m_S^{00}$, *which implies* $\min_{s \in S} \big(m_s^{00} - m_s(z_s)\big) > 0$ *and* $r^* > 0$, *so that* $M(r^*, z_N)$ *is non-empty.*

By definition of $M(r^*, z_N)$ *and because preferences satisfy Strict monotonicity in own consumption, there exists some* z'_N *constructed as in* (1.12) *such that conditions* (1.14) *and* (1.15) *are satisfied. There only remains to show that at* m'_n *condition* (1.16) *is satisfied.*

Because $m'_N \in M(r^*, z_N)$ *we can apply Lemma 2 to get*

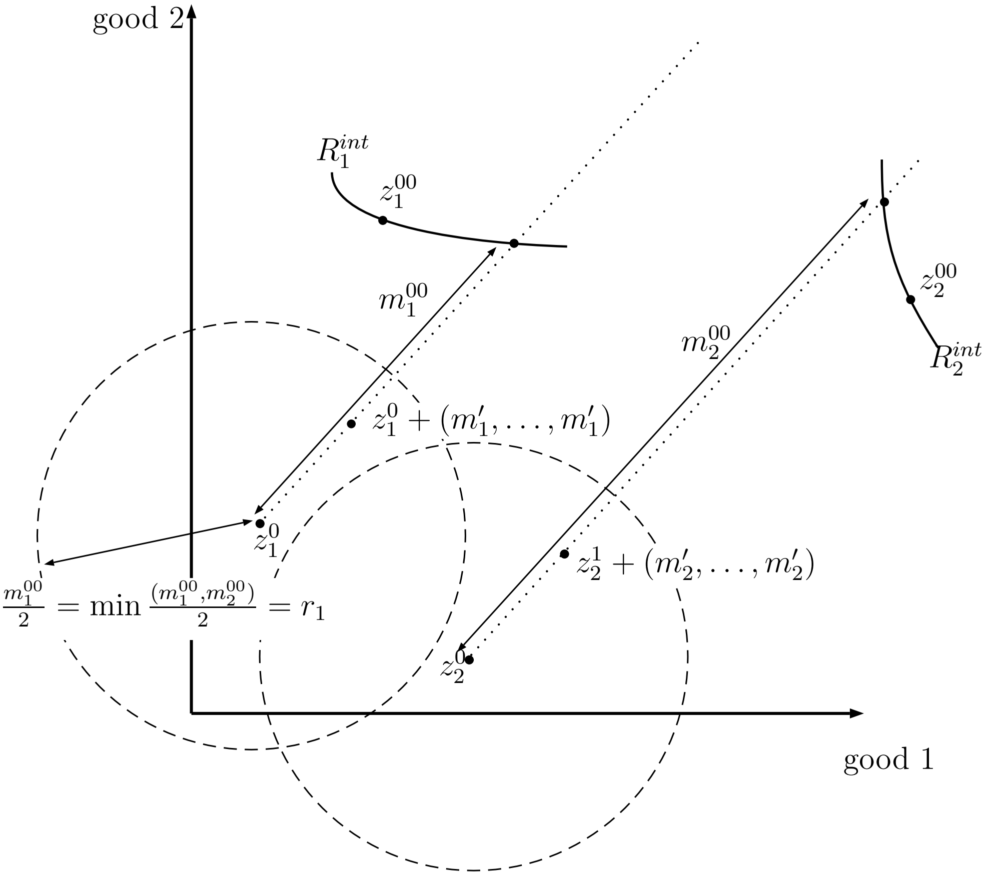

Figure 1.10: Choosing an allocation to satisfy Lemma 3.

$$\begin{aligned}
m'_N &\leq m_N(z_N) + (r^*, \ldots, r^*), \\
m'_S &\leq m_S(z_S) + (r^*, \ldots, r^*), \\
m'_S &\leq m_S(z_S) + \left(\frac{\min_{s \in S} \left(m_s^{00} - m_s(z_s) \right)}{2}, \ldots, \frac{\min_{s \in S} \left(m_s^{00} - m_s(z_s) \right)}{2} \right) \\
& \qquad \textit{[by construction of } r^* \textit{]}, \\
m'_S - m_S(z_S) &\leq \frac{m_S^{00} - m_S(z_S)}{2} \\
m'_S - m_S(z_S) &<< m_S^{00} - m_S(z_S) \qquad \textit{[by } m_S^{00} >> m_S(z_S) \textit{]}, \\
m'_S &<< m_S^{00},
\end{aligned}$$

the desired result.

■

We are now ready to prove the general result.

By construction of z_N^0 and given that preferences satisfy Strict monotonicity in own consumption, $m_S(z_S^0) << m_S(z_S^{00})$. So starting from z_N^0, Lemma 3 applies and we can obtain z_N^* satisfying the conditions (1.14), (1.15) and (1.16) with respect to

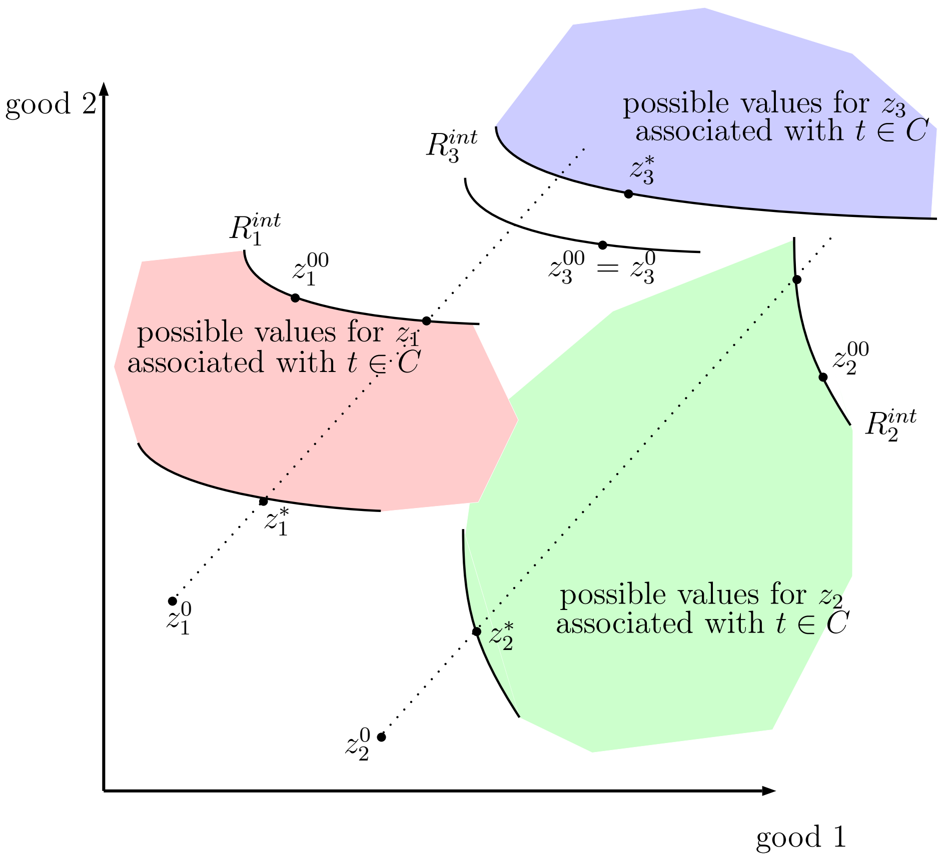

Figure 1.11: Possible allocations associated with $t \in C$, where $S = \{1, 2\}$.

z_N^0. In particular

$$z_N^* \ P_i \ z_N^0, \text{for all } i \in N, \tag{1.17}$$

Notice that z_N^* also satisfies the conditions of Lemma 3. Thus we can apply the lemma again to z_N^* and get a nonempty set of allocations z_N such that

$$\begin{aligned} z_N \ R_i \ z_N^*, & \qquad \text{for all } i \in N, & (1.18) \\ m_i(z_i^*) \leq m_i(z_i), & \qquad \text{for all } i \in N, & (1.19) \\ \text{and } m_j(z_j) \leq m_j^{00}, & \qquad \text{for all } j \in S, & \end{aligned}$$

were we voluntarily turned strict inequalities and preference relations into weak ones. In particular, the last two inequalities imply

$$0 \leq m_j^{00} - m_j(z_j) \leq m_j^{00} - m_j(z_j^*), \qquad \text{for all } j \in S, \tag{1.20}$$

which in turn implies

$$0 \leq \max_{j \in S}(m_j^{00} - m_j(z_j)) \leq \max_{j \in S}(m_j^{00} - m_j(z_j^*)). \tag{1.21}$$

Next we define C, the set of $\max_{j \in S}(m_j^{00} - m_j(z_j)) \in \left[0, \max_{j \in S}(m_j^{00} - m_j(z_j^*))\right]$

which can be obtained via some z_N satisfying conditions (1.18), (1.19) and (1.20). Possibles z_N are illustrated in Figure 16 for the case of 3 agents. Formally

$$
\begin{aligned}
C := \Big\{ & t \in \big[0, \max_{j\in S}(m_j^{00} - m_j(z_j^*))\big] \mid \\
& t = \max_{j\in S}(m_j^{00} - m_j(z_j)) \text{ for some } z_N \in Z \text{ satisfying} \\
& z_N \; R_j \; z_N^*, && \text{for all } i \in N, \\
& m_i(z_i^*) \leq m_i(z_i), && \text{for all } i \in N, \\
\text{and } & 0 \leq m_j^{00} - m_j(z_j) \leq m_j^{00} - m_j(z_j^*), && \text{for all } j \in S \Big\}.
\end{aligned}
$$

The rest of the argument follows the intuition we gave at the beginning of the proof, using the set C to construct the appropriate allocations. We will be interested in $c := \inf C$. We will consider two cases. In case 1, $c = 0$. We will show that given the assumptions on our domain, this means that all agents in S are back to their initial internal utility level m_S^{00}, which will lead to a contradiction.

In case two, $c > 0$, and there will be two subcases. In the subcase 1, we will assume $(m_j^{00} - m_j(z_j^{**})) > 0$ for all $j \in S$ in which case we will be able to apply Lemma 3 again and show that c could not have been the infimum of C. In subcase 2, we will assume that $(m_j^{00} - m_j(z_j^{**})) > 0$ for some subset of $S_1 \subset S$ only. Then we will be able to repeat the whole argument several times up to the point where S_{n^*} brings us back to case 1.

In order to solve those two cases, we will need to associate c with an allocation that everyone weakly prefers to z_N^*. Because c is the infimum of C, and C might not be closed, we have no guaranteed that $c \in C$, and such allocation might not exist. However, we will be able to construct it under the assumptions on our domain. The next part of the proof describes this construction.

By (1.18), (1.19) and (1.20), C is nonempty. Because C is also bounded, c is well-defined. Because c is the infimum of C, there is a sequence $\{c^g\}_{g=1}^{\infty}$ such that for all $g \in \mathbb{N}$, $c^g \in C$, and $c^g \to c$. By definition of C this means that there is a corresponding sequence $\{\tilde{z}_N^g\}_{g=1}^{\infty}$ such that for all $g \in \mathbb{N}$,

$$
\begin{aligned}
& c^g = \max_{j\in S}(m_j^{00} - m_j(\tilde{z}_j^g)), && && (1.22) \\
& \tilde{z}_N^g \; R_j \; z_N^*, && \text{for all } i \in N, && (1.23) \\
& m_i(z_i^*) \leq m_i(\tilde{z}_i^g), && \text{for all } i \in N, && (1.24) \\
\text{and } & 0 \leq m_j^{00} - m_j(\tilde{z}_j^g) \leq m_j^{00} - m_j(z_j^*), && \text{for all } j \in S && (1.25)
\end{aligned}
$$

The next lemma shows that $\{\tilde{z}_N^g\}_{g=1}^{\infty}$ can be turned into a bounded sequence having similar properties for $j \in N$. By the Bolzano-Weierstrass theorem (Rudin, 1976, Theorem 3.6 (b)), we will then be able to find a converging subsequence which will allow us to apply Continuity.

Lemma 4 ($\{\tilde{z}_N^g\}_{g=1}^{\infty}$ can be turned into an equivalent bounded sequence)**.** *Based on $\{\tilde{z}_N^g\}_{g=1}^{\infty}$, we can construct yet another sequence $\{\hat{z}_N^g\}_{g=1}^{\infty}$, such that for all $g \in \mathbb{N}$,*

$$\begin{aligned} &\hat{z}_N^g \; R_j \; \tilde{z}_N^g, && \textit{for all } j \in S, && (1.26)\\ &m_j(\hat{z}_j^g) \;=\; m_j(\tilde{z}_j^g), && \textit{for all } j \in S, && (1.27)\\ &\text{and } \{\hat{z}_N^g\}_{g=1}^{\infty} \text{ is bounded.} && && (1.28) \end{aligned}$$

Proof. The construction goes as follows.

$$\begin{aligned} &\hat{z}_j^g := z_j^0 + (m_j(\tilde{z}_j^g), \ldots, m_j(\tilde{z}_j^g)), && \textit{for all } j \in S, && (1.29)\\ &\hat{z}_k^g := z_k^{00}, && \textit{for all } k \in N\backslash S. && (1.30) \end{aligned}$$

By construction of $\hat{z}_N$ and by definition of m_j, (1.27) is immediate. Then by (1.25) it follows directly that for all $j \in S$ and for all $g \in \mathbb{N}$,

$$0 \le m_j^{00} - m_j(\tilde{z}_j^g) \le m_j^{00} - m_j(z_j^*),$$
$$0 \le m_j(z_j^*) \le m_j(\tilde{z}_j^g).$$

But then by construction $\{\hat{z}_S^g\}_{g=1}^{\infty}$ is bounded. Then, because $\{\hat{z}_{N\backslash S}^g\}_{g=1}^{\infty}$ is constant, $\{\hat{z}_N^g\}_{g=1}^{\infty}$ is bounded.

We now prove (3.28). By construction we have

$$\hat{z}_j^g \; I_j^{int} \; \tilde{z}_j^g, \qquad \textit{for all } j \in S. \tag{1.31}$$

By Separability, this implies

$$(\hat{z}_S^g, \tilde{z}_{N\backslash S}^g) \; I_j \; \tilde{z}_N^g, \qquad \textit{for all } j \in S, \tag{1.32}$$

Now by (1.24), given that z_N^* satisfies (1.16) with respect to z_N^0 and that for all $k \in N\backslash S$, $z_k^0 = z_k^{00}$ by assumption, we have

$$m_k(\tilde{z}_k^g) > m_k^{00}, \qquad \textit{for all } k \in N\backslash S. \tag{1.33}$$

Therefore by No altruism and Well-being externality

$$(\hat{z}_S^g, \hat{z}_{N\backslash S}^g) \; R_j \; (\hat{z}_S^g, \tilde{z}_{N\backslash S}^g), \qquad \textit{for all } j \in S, \tag{1.34}$$

which by transitivity implies

$$(\hat{z}_S^g, \hat{z}_{N\backslash S}^g) \; R_j \; \tilde{z}_N^g, \qquad \textit{for all } j \in S, \tag{1.35}$$

the desired result.

■

Now by transitivity and (1.23), Lemma 4 implies that for all $g \in \mathbb{N}$,

$$\hat{z}_N^g \; R_j \; z_N^*, \qquad \text{for all } j \in S, \tag{1.36}$$

where $\{\hat{z}_N^g\}_{g=1}^{\infty}$ is as constructed in the lemma. Because $\{\hat{z}_{N\backslash S}^g\}_{g=1}^{\infty}$ is bounded, by the Bolzano-Weierstrass theorem, the sequence has a converging subsequence, say

$\{\hat{z}^{t(g)}_{N\setminus S}\}_{g=1}^{\infty}$. Let z^{**}_N be the limit of $\{\hat{z}^{t(g)}_{N\setminus S}\}_{g=1}^{\infty}$. By Continuity,

$$z^{**}_N \; R_j \; z^{*}_N, \qquad \text{for all } j \in S. \tag{1.37}$$

By transitivity, (1.17), and given the contradiction assumption (1.7), we then have

$$z^{**}_N \; P_j \; z^{00}_N, \qquad \text{for all } j \in S. \tag{1.38}$$

By (1.25) and (1.27) we have that for all $g \in \mathbb{N}$

$$0 \le m^{00}_j - m_j(\hat{z}^{t(g)}_j), \qquad \text{for all } j \in S, \tag{1.39}$$

Thus by continuity of m_j we get

$$0 \le m^{00}_j - m_j(z^{**}_j), \qquad \text{for all } j \in S, \tag{1.40}$$

Also, by (1.22) and (1.27), for all $g \in \mathbb{N}$,

$$c^{t(g)} = \max_{j\in S}(m^{00}_j - m_j(z^{t(g)}_j)) = \max_{j\in S}(m^{00}_j - m_j(\hat{z}^{t(g)}_j)),$$

So by Continuity of m

$$c = \max_{j\in S}(m^{00}_j - m_j(z^{**}_j)),$$

Remember that by definition of C, we have $c \in \left[0, \max_{j\in S}(m^{00}_j - m_j(z^{*}_j))\right]$. Finally notice that because $\hat{z}^{t(g)}_k = z^{00}_k$ for all $k \in N\setminus S$ and every $g \in \mathbb{N}$, we also have

$$z^{**}_k = z^{00}_k, \qquad \text{for all } k \in N\setminus S. \tag{1.41}$$

We are now ready to study the two cases mentioned above.

Case 1 : $\max_{j\in S}(m^{00}_j - m_j(z^{**}_j)) = 0$**.** This is equivalent to

$$0 \ge m^{00}_j - m_j(z^{**}_j), \qquad \text{for all } j \in S. \tag{1.42}$$

By (1.40) and (1.42) we have

$$0 = m^{00}_j - m_j(z^{**}_j), \qquad \text{for all } j \in S. \tag{1.43}$$

By Well-being externality, this means

$$(z^{00}_S, z^{**}_{N\setminus S}) \; I_j \; z^{**}_N, \qquad \text{for all } j \in S. \tag{1.44}$$

But by (1.41),

$$(z^{00}_S, z^{**}_{N\setminus S}) = (z^{00}_S, z^{00}_{N\setminus S}) = z^{00}_N. \tag{1.45}$$

Thus (1.44) can be rewritten as

$$z^{00}_N \; I_j \; z^{**}_N, \qquad \text{for all } j \in S. \tag{1.46}$$

contradicting (1.38).

Case 2 : $\max_{j\in S}(m_j^{00} - m_j(z_j^{**})) > 0$**.** There are two subcases.

Subcase 1 : $(m_j^{00} - m_j(z_j^{**})) > 0$ **for all** $j \in S$**.** Then notice that the assumptions of Lemma 3 hold at z_N^{**}. So we can apply the lemma once again and obtain an allocation z_N^{***} which is associated with some $r^{***} \in C$ such that $r^{***} < c$, contradicting the fact that c is the infimum of C.

Subcase 2 : there exists a nonempty $\tilde{S}_1 \subset S$ **with** $(m_j^{00} - m_j(z_j^{**})) = 0$ **for all** $j \in \tilde{S}_1$**.**

Slightly abusing the notation, let $\tilde{S}_1$ be the largest such set. Then for any $j \in S^1 := S\backslash\tilde{S}_1$, $(m_j^{00} - m_j(z_j^{**})) > 0$. Thus we can repeat the former steps.

By Well-being externality, we have

$$(z_{S_1}^{**}, z_{\tilde{S}_1}^{00}, z_{N\backslash S}^{**})\ I_h\ z_N^{**}, \qquad \text{for all } h \in S. \tag{1.47}$$

Thus by (1.33), Well-being externality, and No altruism we have

$$(z_{S_1}^{**}, z_{\tilde{S}_1}^{00}, z_{N\backslash S}^{00})\ R_h\ (z_{S_1}^{**}, z_{\tilde{S}_1}^{00}, z_{N\backslash S}^{**}), \qquad \text{for all } h \in S_1. \tag{1.48}$$

which by transitivity yields

$$(z_{S_1}^{**}, z_{N\backslash S_1}^{00})\ R_h\ z_N^{**}, \qquad \text{for all } h \in S_1, \tag{1.49}$$

and

$$(z_{S_1}^{**}, z_{N\backslash S_1}^{00})\ P_h\ z_N^{00}, \qquad \text{for all } h \in S_1. \tag{1.50}$$

Notice that this brings us back to an allocation $(z_{S_1}^{**}, z_{N\backslash S_1}^{00})$ very similar to z_N^0, except that the relevant set of agents is now $S_1 \subset S$ instead of S. Starting from $(z_{S_1}^{**}, z_{N\backslash S_1}^{00})$ we can repeat the whole argument as many times as we want. Every time we do so, we get smaller and smaller sets $S_n \subset \cdots \subset S_1 \subset S$.

Because N is finite, either we reach subcase 1 directly for some S_{n^*} and get a contradiction, or there is some $S_{n^{**}}$ with a single agent. But if $S_{n^{**}}$ contains a single agent, we again reach Case 1 and get a contradiction. Hence we are done.

Chapter 2

A new index combining the absolute and relative aspects of income poverty: theory and application

2.1 Introduction

Income poverty reduction is a major political objective, both at national and international levels. In the past decade, policy makers such as the EU Commission or the World Bank have adopted quantified poverty reduction targets.[1] These targets are based on income poverty measures, which are composed of two elements: a poverty line and an index (Sen, 1976). A poverty line specifies the income threshold below which individuals are considered to be poor. An index aggregates the poverty of all individuals in a society and, hence, allows us to compare poverty in different societies.

There exist two central approaches for measuring income poverty, absolute poverty and relative poverty. They differ in the type of poverty line used. An absolute line has its income threshold independent of the standard of living whereas a relative line's income threshold evolves as a constant fraction of the standard of living. These two types of lines aim at capturing different deprivations. On the one hand, absolute poverty refers to the idea of *subsistence*. An individual is absolutely poor if her income is not sufficient to satisfy several of her basic needs, such as being sufficiently nourished. In a first approximation, the real cost of subsistence is absolute as it does not depend on standards of living. For example, 100 grams of rice contain the same amount of calories in New-York or in New-Delhi. On the other hand, relative poverty refers to the ideas of *social participation* or *inclusion*. An individual is relatively poor if her income is not sufficient to engage in the everyday life of her society (Townsend, 1979; Sen, 1983). The real cost of not being excluded from social participation is relative as it depends on standards of living. The archetypical example is that of the linen shirt (Smith, 1776). Adam Smith observed that in the England of his time people would be too ashamed to appear in public without wearing a

[1] See World Bank (2015) or European Commission (2015).

linen shirt, which he argued was not the case in the Roman Empire that had a lower standard of living.[2]

Many policy makers aim at reducing both the absolute and relative poverties. These two objectives appear for example in the poverty reduction target of the EU Commission or in the new twin goals of the World Bank.[3] Against absolute poverty policy makers implement pro-growth policies, which typically reward efforts at the potential cost of increasing inequalities. Increasing a poor individual's income improves her absolute poverty but increasing the inequality she experiences worsens her relative poverty. Against relative poverty policy makers implement redistributive policies, some of which may distort incentives. Of course, not all policies induce a trade-off between growth and equality. Nevertheless, one policy seldom dominates all the altenative policies in both dimensions.

As the two objectives are not always aligned, policy makers must regularly arbitrate between them. Trading-off absolute and relative poverty amounts to answering the following question: when does *unequal growth* alleviates income poverty? A country experiences unequal growth if its economic growth goes along with an increase in income inequality. That is, all individuals get more resources but the additional resources go disproportionately more to the middle class and the rich than to the poor.

One serious difficulty is that the two measurement approaches make opposite extreme judgments on unequal growth. Hence, they evaluate very differently the merits of development programs leading to unequal growth. On the one hand, absolute measures evaluate growth positively, regardless of its distribution. On the other hand, relative measures judge positively any reduction in the inequality experienced by the poor, regardless of the poor's income level. Clearly, neither absolute measures nor relative measures are able to make this trade-off. Measuring both forms of poverty in parallel does not solve the issue since, more often than not, the two approaches yield opposite conclusions.[4]

This paper proposes a new way to measure poverty that combines the absolute and relative aspects of income poverty. Previous attempts to develop such a measure

[2]The normative foundations for taking a relativist approach in poverty measurement are reviewed in Ravallion (2008). For instance, Sen (1983) made the case that an absolute level in the space of capabilities translates into a relative level in the space of resources. Townsend (1979) discussed how individuals not having the resources for obtaining the living conditions that are widely encouraged in their society would be excluded from ordinary living patterns, customs and activities. Runciman (1966) pointed out that the comparison of own income with incomes of better-off individuals creates a feeling of deprivation.

[3]In its EU2020 strategy, the EU Commission targets to reduce by 20 millions the number of individuals that are *at risk of poverty or social exclusion* (AROPE). The AROPE individuals are *inter alia* those individuals that are *at risk of poverty* (relative poverty) or are *severely materially deprived* (absolute poverty). In 2013, the World Bank committed itself to twin goals: eliminating extreme poverty (absolute poverty) and boosting shared prosperity (relative poverty). The second objective has a clear relative flavor since it is defined as raising the living standards of the bottom 40% of individuals *in any given country.*

[4]A common practice is to use absolute measures in low- and middle-income countries and relative measures in high-income countries. Official national poverty definitions mostly follow this practice (Ravallion, 2012) that leads to extreme judgments as explained above.

followed two different routes. One route proposes to measure both forms of poverty in parallel before looking for a way to aggregate them (Atkinson and Bourguignon, 2001; Anderson and Esposito, 2013). Unfortunately, this approach is confronted to several difficulties, including double counting issues. The other route aims at developing a single measure based on a poverty line making the trade-off between the absolute and relative aspects of income. The most influential proposals of such *endogenous* lines are the hybrid lines (Foster, 1998) and the weakly relative lines (Ravallion and Chen, 2011). So far, this second route has mostly focused on defining new poverty lines. Surprisingly, indices to use in combination with an endogenous line have not been rigorously studied. In empirical applications (Chen and Ravallion, 2013), the default practice is to use an endogenous line in combination with an index derived for absolute lines, such as the very popular Foster-Greer-Thorbecke (FGT) indices (Foster et al., 1984).

As shown in this paper, there are two limitations associated with measures combining an endogenous line and an index derived for absolute lines. First, indices derived for absolute lines loose some of their desirable properties when combined with endogenous lines. Second, the endogenous measures obtained by this practice weigh the absolute and relative aspects of income poverty in a questionable way. They may consider that absolutely poor individuals in low-income countries are *less poor* than relatively poor individuals in middle- and high-income countries. The problem is so serious that these endogenous measures may conclude that there is more poverty in the latter countries than in low-income countries. In the application, measures composed of an endogenous line and an FGT index deem Brazil equally or more poor than Ivory Coast in 2010. Even if income inequality was larger in Brazil than in Ivory Coast, such judgment could be seriously questioned given that mean income in Brazil was more than four times larger than that of Ivory Coast. Moreover, 22.7 % of individuals in Ivory Coast lived on less than 1.25 $ a day – the World Bank's threshold for extreme poverty (Ravallion et al., 2009) – but only 5.4% in Brazil.

Why do measures combining an endogenous line with an FGT index yield this debatable conclusion? FGT indices implicitly attribute to each individual a value of *individual poverty* that depends only on her *normalized income*, i.e. her income divided by the income threshold in her society. In 2010, an individual living on 1 $ a day in Ivory Coast has the same normalized income as an individual living on 3.6 $ a day in Brazil for the weakly relative line used by Chen and Ravallion (2013). As a result, FGT indices attribute to both the same individual poverty. This conclusion ignores that, unlike the latter, the individual in Ivory Coast is below the threshold for extreme poverty. Being extremely poor is not reflected in normalized incomes. Hence, an extremely poor individual in Ivory Coast can be deemed less poor than a non-extremely poor individual in Brazil. This problem is not limited to indices based on normalized incomes but is rather pervasive. It also affects indices based on absolute gaps, i.e. the distance between the threshold and the individual income.

This paper proposes a new index combining the absolute and relative aspects of income poverty. In order to avoid the problem faced by standard indices, I depart from individual poverty comparisons based on normalized incomes. To begin with, I define an absolute poverty threshold, which in the application is fixed at 1.25 $ a day. Below this subsistence threshold, an individual is deemed absolutely poor and her individual poverty does not depend on the standard of living in her society. For instance, two individuals living with 1.25 $ a day in Ivory Coast and Brazil contribute identically to poverty in their respective countries. Then, I define the endogenous poverty line above the absolute threshold. An individual above the absolute threshold but below the endogenous line is deemed relatively poor. Her individual poverty depends on the standard of living in her society. In the application, an individual living on 2 $ a day in Ivory Coast, where the mean is 3 $ a day, contributes identically to poverty as an individual living on 6.8 $ a day in Brazil, where the mean is 13.8 $ a day.

More generally, I formalize the comparison of individual poverties across societies having different standards of living by defining the concept of *equivalence ordering*. In a nutshell, two individuals that are attributed equal individual poverties are on the same equivalence curve. I constrain equivalence curves below the absolute threshold to be independent of standards of living. In contrast, the equivalence curves above the absolute threshold may evolve with standards of living. The constraints I impose on equivalence curves imply that absolutely poor agents are always considered poorer than relatively poor agents. This judgment is in line with largely shared intuitions, as appeared from questionnaire studies run all over the world by Corazzini et al. (2011).

This paper has two main theoretical results. First, I characterize a family of additive indices based on mean-sensitive endogenous poverty lines. In other words, I identify the set of properties defining a family of indices based on poverty lines sensitive to mean income. This is the first characterization of indices based on non-absolute lines. This result extends the characterization of additive indices of Foster and Shorrocks (1991) to non-absolute lines. Then, I investigate which members of this additive family satisfy compelling properties. To do so, I define an extended family of FGT indices based on equivalence orderings meeting the constraints mentioned above. This family depends on two parameters, one of which is the poverty aversion parameter. The second result shows that a unique member of this extended FGT family satisfies two basic properties. One property is classical and requires that a progressive transfer between two poor individuals does not increase poverty. The other property is new and specific to indices based on endogenous lines. It requires that destroying part of the income of a poor individual does not reduce poverty. This property excludes all values of poverty aversion except the one associated to the Poverty Gap Ratio.

The index characterized is new and inherits the properties of its underlying equivalence ordering. That is, absolutely poor individuals are distinguished from relatively poor individuals and the former are always considered poorer than the latter. Being additive, the new index is decomposable between the respective contributions of absolutely and relatively poor individuals. This last feature simplifies the analysis of the evolution of poverty and its communication.

Finally, a poverty measure based on the new index is applied to World Bank data. This application illustrates that the judgments obtained from the new measure are more in line with general intuitions than those obtained with standard measures. For instance, the new measure deems Brazil less poor than Ivory Coast. In a second step, the new measure is used to assess the evolution of poverty in several countries that experienced unequal growth. Urban China constitutes a proeminent example because it experienced over the period 1990 – 2010 a strong growth together with a sharp increase in inequality. The new measure concludes that poverty in urban China was reduced by about 75% over this period. By decomposing the measure, one can see that this improvement almost entirely rests on the drastic reduction in absolute poverty. Absolute poverty accounted for about two-third of income poverty in 1990, but less than 10% in 2010. This shows that if the main issue in urban China was absolute poverty in 1990, it has become relative poverty in 2010. Studying different countries shows that the measure may yield different judgments on unequal growth. Over the period 1990 – 2010, income poverty did not change in Mexico as the reduction in absolute poverty was compensated by the increase in relative poverty. Over the period 1996 – 2010, unequal growth has lead to an increase in poverty in Hungary where the impact on relative poverty was dominant. In general, whether unequal growth reduces the poverty measure or not depends on the initial importance of absolute poverty.

The paper is organized as follows. A literature review is presented in Section 2.2. The framework and the characterization of the additive family are presented in Section 2.3. The index proposed is derived and discussed in Section 2.4. The robustness of the results is investigated in Section 2.5. Other income standards than the mean are discussed in Section 2.6. The empirical illustration is presented in Section 2.7. I conclude in Section 2.8. All proofs are relegated in the Appendix.

2.2 Literature review

I review in this section the literature on income poverty measurement. More specifically, I present the poverty measures that are popular in empirical applications and I emphasize their limitations when comparing societies with different standards of living.

The objective of the literature on income poverty measurement is to rank income distributions with respect to the poverty they contain. I divide all existing measures between those based on absolute lines and those based on endogenous lines. Initially, most contributions were concerned with indices based on absolute lines. This early literature on absolute measures is nicely reviewed in Zheng (1997).

2.2.1 Absolute measures

Absolute measures are measures based on absolute lines. A poverty line is absolute if its income threshold does not evolve with standards of living. I make two remarks on absolute lines in order to avoid any confusion. First, the threshold of an absolute line is constant *in real terms*. To be sure, all incomes in this paper are expressed in real terms. The threshold of an absolute line is often defined by the cost of a particular bundle of goods. The line is then “anchored” in that bundle. This does not prevent the *nominal* threshold of the absolute line to evolve over time with inflation or to vary from one country to another as a function of purchasing power. Second, the bundle of goods “anchoring” an absolute line can potentially capture *both* subsistence and social participation for a given society at a given time. Absolute measures can therefore account for both functionings, but only when comparing two societies having the *same* standards of living.

I present here the notation necessary for exposing the relevant results in the literature on absolute measures. Let an income distribution $y := (y_1, \dots, y_n)$ be a list of non-negative incomes sorted in non-decreasing order ($y_1 \leq \cdots \leq y_n$). Absolute poverty lines are defined by a constant threshold $z^* \in \mathbb{R}_{++}$.[5] Agent i qualifies as poor if $y_i < z^*$. The objective is to rank all distributions in a set Y. Let a poverty index be a real-valued function $P : Y \times \mathbb{R}_{++} \to \mathbb{R}$ representing the complete ranking on Y. For any two $y, y' \in Y$, there is strictly more poverty in y than in y' if $P(y, z^*) > P(y', z^*)$, and weakly more if $P(y, z^*) \geq P(y', z^*)$. The number of poor agents is denoted $q(y)$ or simply q when no confusion is possible.[6] Similarly, the number of agents in y is denoted $n(y)$ or n. Since income distributions are ordered, if $i \leq q$ then agent i is poor.

A central result is the characterization of additive indices. Given an absolute line, any index satisfying five basic properties must be ordinally equivalent to an additive index (Foster and Shorrocks, 1991):

$$P(y, z^*) := \frac{1}{n} \sum_{i=1}^{n} d(y_i), \tag{2.1}$$

where function $d : \mathbb{R}_+ \to [0, 1]$ is non-increasing in y_i and returns zero for all incomes above z^*. The value returned by the function d can be interpreted as the individual poverty associated to earning income y_i. This individual poverty only depends on own income (and z^*). An additive index can be interpreted as the average individual poverty in the distribution.

This family is very broad as very few restrictions are imposed on the function d. The Foster-Greer-Thorbecke (FGT) subfamily proposes an exponential expression for the function d (Foster et al., 1984):

$$P^{FGT}(y, z^*) := \frac{1}{n} \sum_{i=1}^{q} \left(\frac{z^* - y_i}{z^*} \right)^{\alpha}. \tag{2.2}$$

The FGT family has a unique parameter $\alpha \in [0, \infty)$, which can be interpreted as poverty aversion. The larger α, the higher is the priority given by the index to agents at the bottom of the income distribution. This family allows for a very wide variety of judgments and admits the Head-Count Ratio (HC) and the Poverty Gap

[5] $\mathbb{R}_+$ denotes the set of non-negative reals and $\mathbb{R}_{++}$ the set of strictly positive reals.

[6] Although it is not explicit in its notation, q depends on z^*.

Ratio (PGR) as particular cases:

$$HC(y, z^*) := \frac{q}{n} \qquad \text{corresponds to } \alpha = 0,$$

$$PGR(y, z^*) := \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z^* - y_i}{z^*}\right) \qquad \text{corresponds to } \alpha = 1.$$

Virtually all empirical applications use a poverty measure based on an index in the FGT family. Many other absolute indices have been proposed, including those of Kakwani (1980), Chakravarty (1983) or Duclos and Gregoire (2002).

Limitation of absolute measures

Absolute measures are not well-suited for evaluating the impact on poverty of unequal growth. Growth increases the standard of living, which in turns raises the cost of social participation.[7] The shortcoming of absolute measures is that they completely ignore these social participation effects. In a nutshell, they ignore the relative aspect of income poverty.

Table 2.1 presents an example illustrating the problem. This example compares the income distributions of two societies A and B, each populated by two agents, one poor and one non-poor. To fix ideas, assume incomes are given in \$ a day. I assume that the subsistence threshold, denoted by z^a, is \$1.25 a day. This subsistence threshold is smaller than the income threshold z^* defining income poverty in this example. Distribution B dominates distribution A but all extra resources go to the non-poor individual, except for an epsilon. Any absolute measure concludes there is less poverty in society B than in A. This conclusion is debatable for small epsilon. If the poor individual has more income in B than in A, she is worse off in A than in B in the two-dimensional space relevant for poverty evalution: subsistence and social participation. Indeed, the poor individual is above the subsistence threshold in both societies but has more difficulties to participate in society B than in society A.[8]

Table 2.1: Absolute measures ignore social participation effects.

	y_1	y_2	z^a	z^*
Society A	3	15	1.25	5
Society B	$3+\epsilon$	100	1.25	5

The problem illustrated in Table 2.1 results from the axiom of *Focus*. This axiom is satisfied by all indices derived for absolute lines. *Focus* encapsulates the key property distinguishing poverty indices from inequality indices, namely that

[7] This assumption is in line with evidence provided by national poverty thresholds. In purchasing power parity, national income thresholds tend to increase with standards of living (Ravallion, 2012).

[8] Empirical Social Choice has shown from questionnaire experiments that resources dominance is far from being unanimously accepted by respondents as a sufficient normative criteria for concluding that one distribution is better than another one (Gaertner and Schokkaert, 2012). A tentative explanation put forward by this literature is that respondents consider other-regarding feelings and hence would agree with dominance in the space of utilities, but not in the space of resources.

poverty indices are only concerned with the fate of poor agents. Formally, *Focus* requires the index not to be affected by the income of non-poor agents.

Social ordering axiom 11 (*Focus*).
For all $y, y' \in Y$ *and* $z^* \in \mathbb{R}_{++}$, *if* $n(y) = n(y')$, $q(y) = q(y')$ *and* $y_i = y'_i$ *for all* $i \leq q(y)$, *then* $P(y, z^*) = P(y', z^*)$.

A corollary of *Focus* is that a distribution has the same poverty as its associated censored distribution, obtained by replacing the income of all non-poor agents by the income threshold z^*. This axiom, together with a monotonic property, implies that if a censored distribution first-order stochastically dominates another, then it has less poverty, excluding social participation effects.

Social participation effects can be accounted for in the identification of poverty by the use of endogenous lines. However, measures obtained by combining endogenous lines with indices derived for absolute lines fail to give a minimal priority to subsistence over social participation.

2.2.2 Endogenous measures

Endogenous measures are measures based on endogenous lines. The income threshold of an endogenous line may evolve with standards of living. In practice, the threshold is endogenously determined by the mediation of an income standard like mean or median income.

Relative lines are the most famous example of endogenous lines. The threshold of a relative line evolves as a constant fraction of the income standard. Relative lines are widely used in developed countries. For example, the "At Risk of Poverty" measure of the European Commission is based on a relative line. Nevertheless, relative lines have been heavily criticized, mainly on two grounds (Ravallion and Chen, 2011). First, their threshold goes to zero in low-income countries, making clear that subsistence is not accounted for. Second, relative lines are based on a rather extreme view on social participation. Indeed, any equi-proportionate growth does not get any individual out of poverty because the threshold is multiplied by the same factor as the individual incomes. Poverty measures based on an FGT index together with a relative line are unaffected by equi-proportionate growth. In that sense, these relative measures ignore absolute gains and losses.

Given the shortcomings of relative measures, a literature emerged with the ambition to balance the absolute and relative aspects of income poverty, albeit most efforts concentrated on identification. Two main routes have been proposed for endogenous identification.

The first route consists in using two different lines for identification, one absolute and one relative (Atkinson and Bourguignon, 2001). The absolute line captures subsistence, referred to as "absolute poverty", and the relative line captures social participation, referred to as "relative poverty". As the relative line's threshold is larger than that of the absolute line in high-income countries and smaller in low-income countries, the two lines cross. As a consequence, some individuals in low-income countries can be deemed absolutely poor but not relatively poor. If this route proposes a meaningful way of identifying the poor, the construction of a good index based on two lines has proved very difficult. The first solution is to construct two measures, one based on the absolute line and the other on the relative line. In order to judge unequal growth, the two measures need to be aggregated. Atkinson and Bourguignon (2001) suggest to consider the two measures in lexicographic order, as they judge subsistence to be a more serious component of poverty than social participation. Lexicographic aggregation unfortunately makes the relative measure almost irrelevant in poverty judgments. Another possibility is to weight the two measures (Anderson and Esposito, 2013). The second solution is to derive a unique index by aggregating the income gaps with respect to each of the two lines (Atkinson and Bourguignon, 2001). Unfortunately, this raises a problem of double counting for individuals that are both absolutely and relatively poor. So far, a characterization of the properties of an index based on two lines remains missing. As a result, there is no guarantee that measures based on two lines give a minimal priority to subsistence over social participation.

The second route consists in using a unique endogenous line balancing the absolute and relative aspects of income poverty. Foster (1998) proposes *hybrid* lines that feature a constant income elasticity ρ.[9] This income elasticity can be interpreted as the extent to which poor individuals should share the benefits of economic growth. Absolute lines have an income elasticity of zero and relative lines have an income elasticity of one, representing two extreme views on this parameter.[10] A different proposal by Ravallion and Chen (2011) suggests using *weakly relative* lines, whose income elasticity is zero for low-income countries and then increases with standards of living, tending ultimately to a value of one. Both hybrid and weakly relative lines are interesting proposals for identifying the poor. Unfortunately, a characterization of the properties of an index based on a unique endogenous line remains missing. So far, all endogenous measures used in empirical applications are based on FGT indices, which are characterized for absolute lines (see for example Chen and Ravallion (2013)). This is problematic as those indices loose some of their desirable properties when combined with endogenous lines.

Limitations of endogenous measures

Current endogenous measures are not well-suited for evaluating the impact on poverty of unequal growth. They give no priority to subsistence over social participation. The judgment that subsistence should be given priority over social participation is not only the intuition of experts like Atkinson and Bourguignon (2001) but appears to be largely shared as shown in questionnaire studies conducted in different parts of the World (Corazzini et al., 2011).

Current endogenous measures consider that some individuals whose income is below the subsistence level are less poor than other individuals living in a richer

[9]For a given income standard, letting z_a be the threshold of an absolute line and z_r be the threshold of a relative line, the hybrid threshold is given by $z_h = z_a^{\rho} z_r^{1-\rho}$.

[10]Madden (2000) estimates empirically an upper-bound for the value of this parameter using Irish data.

society but whose income is above subsistence level. Notice that I make a pairwise comparison of individuals living in different societies. Such a comparison cannot be performed using a standard poverty axiom because axioms compare two distributions, not two individuals. The issue is nevertheless extremely serious as it often leads endogenous measures to conclude that there is more poverty in middle and high-income societies than in low-income societies. In other words, endogenous measures often conclude that growth increase poverty even if many individuals were brought above the subsistence level.

Table 2.2 presents an example illustrating the problem. Assume for simplicity that the endogenous measure uses an FGT index P based on a relative line z whose threshold is defined as 50% of the mean income (denoted $\overline{y}$). For example, index P could be the HC or the PGR. The *normalized income gap* of a poor individual is defined to be one minus her normalized income:

$$g\big(y_i, z(\overline{y})\big) := \frac{z(\overline{y}) - y_i}{z(\overline{y})} = 1 - \frac{y_i}{z(\overline{y})}. \tag{2.3}$$

FGT indices are then simply the average of normalized income gaps taken to the power α. The example in Table 2.2 compares the income distributions of two societies C and D, each populated by two agents, on poor and one non-poor. The poor individual in society C has income below the subsistence threshold, i.e. below \$1.25 a day, whereas the poor individual in society D has income above the subsistence threshold. Nevertheless, the poor individual in society D has a larger normalized income gap, implying the endogenous measure concludes there is more poverty in D than in C.[11] Observe that for any other monotonic endogenous line, another example featuring the same issue can be constructed.

Table 2.2: Endogenous measures give no priority to subsistence (\$1.25 a day).

	y_1	y_2	z^a	$z(\overline{y})$	$g(y_1, z(\overline{y}))$	$P(y)$
Society C	1	5	1.25	1.5	$\frac{1}{3}$	$\frac{1}{2}\left(\frac{1}{3}\right)^\alpha$
Society D	1.5	10.5	1.25	3	$\frac{1}{2}$	$\frac{1}{2}\left(\frac{1}{2}\right)^\alpha$

The problem illustrated in Table 2.2 results from the axiom of *Scale Invariance*. This axiom is satisfied by virtually all endogenous measures. This axiom is often defended on the grounds that it renders the currency units in which income is measured irrelevant. Formally, *Scale Invariance* requires the index not to be affected when the income of all agents are multiplied by the same factor as the income threshold.

Social ordering axiom 12 (*Scale Invariance*).
For all $y \in Y$ and $\lambda > 0$, $P(y, z^*) = P(\lambda y, \lambda z^*)$.

A corollary of *Scale Invariance* is that a censored distribution has the same poverty as its associated normalized gaps censored distribution, obtained by replac-

[11] The HC concludes that there is equivalent poverty in both societies. A conclusion that is also questionable.

ing all censored incomes by their normalized gaps. Therefore, whether or not an individual has income below the subsistence level is irrelevant. The fact that *Scale Invariance* imposes more than just the irrelevance of currency units has already been emphasized by Zheng (2007). This author derives indices satisfying a weaker property than *Scale Invariance*. His approach is nevertheless only concerned with indices based on absolute lines (his indices satisfy *Focus*).

The empirical illustration provides examples of such debatable judgments by current endogenous indices, for example when comparing Brazil with Ivory Coast in 2010, as explained in the Introduction.[12]

I showed that standard measures provide very counter-intuitive judgments when comparing income poverty between societies with different standards of living. There is a need for a new measure that can provide sound judgments on unequal growth. Clearly, no absolute measure can account for social participation effects. Therefore, I study indices based on a unique endogenous line. As emphasized above, there exists no characterization of any such index. In the next section, I derive an additive family of indices based on endogenous lines.

2.3 Additive indices based on endogenous lines

I describe in this section the characterization of a family of additive indices based on endogenous lines. The family is based on the concept of an equivalence ordering. The presentation of this new object requires the introduction of additional notation.

2.3.1 Notations and basic restrictions

The notation is a slight modification of the notation used for absolute measures, presented in section 2.2. Mean income $\overline{y} := \frac{\sum y_i}{n}$ is the income standard capturing standards of living. This choice and the robustness of the results for other income standards are discussed in Section 2.6. I refer to y_i as the *absolute situation* of agent i and $\frac{y_i}{\overline{y}}$ as her *relative situation*.

An endogenous poverty line is defined by its associated threshold function $z : \mathbb{R}_{++} \to \mathbb{R}_{++}$ specifying the income threshold $z(\overline{y})$ associated to $\overline{y}$. Agent i qualifies as poor if $y_i < z(\overline{y})$. Letting $N := \{n \in \mathbb{N} | n \geq 3\}$, the set of income distributions considered is

$$Y := \left\{ y \in \mathbb{R}_+^N \mid \overline{y} > 0 \text{ and } y_n \geq z(\overline{y}) \right\}.$$

For technical reasons, this set excludes lists of zeros and any distribution for which all agents are poor. These two restrictions are arguably rather mild.

In order to keep the notation minimal, the notation P for poverty indices features the income distribution as its unique argument, suppressing its dependence on the line z. A poverty index is therefore a real valued function $P : Y \to \mathbb{R}$. For any two

[12]From a theoretical perspective, societies having different standards of living can be either the same country at different points in time or two different countries at the same point in time.

$y, y' \in Y$, there is strictly more poverty in y than in y' if $P(y) > P(y')$, and weakly more if $P(y) \geq P(y')$.

Endogenous lines

This research does not provide a guide for the selection of an endogenous line. The endogenous line is assumed to be exogenously given. How can a practitioner select a good line? Ideally, for each value of mean income, the threshold function returns the minimal cost of a bundle of goods sufficient to secure subsistence and social participation.[13] In practice, such a line could be regressed on the costs of a set of reference bundles, each bundle constructed for a different country (the sample should cover countries with different standards of living). A more pragmatic choice is to select an hybrid or a weakly relative line. Most important is that the line makes sense in the societies that are being compared.

The selection of an endogenous line is a normative choice. Therefore, this selection involves some arbitrariness. An important remark is that selecting an endogenous line does not involve more arbitrariness than selecting an absolute line. There are of course fewer parameters to fix when opting for an absolute line, but this is precisely because absolute lines assume the income threshold to be constant. This assumption is as arbitrary as selecting a positive slope for the line. What is more, this assumption implies that the associated poverty measure ignores completely social participation effects, as illustrated in Table 2.1.

For the results to hold, the endogenous line must meet two mild restrictions, besides being continuous. Possibility of Poverty Eradication requires the existence of an income level that, if earned by all agents, makes all agents non-poor.

EL restriction 1 (Possibility of Poverty Eradication)**.**
There exists $g > 0$ *such that* $g \geq z(g)$.

The restriction Slope Less than One requires that the slope of z at mean income $\overline{y}$, denoted $s(\overline{y})$, is never larger than one.[14] This restriction implies that if an agent is not poor in an initial distribution and if her income and mean income increase by the same amount, this agent cannot be considered poor in the new distribution.

EL restriction 2 (Slope Less than One)**.**
For all $\overline{y} > 0$ *we have* $s(\overline{y}) \leq 1$.

Together, the two restrictions imply there exists a minimal value of mean income above which poverty-free income distributions always exist.

The presentation of the results is simplified by the introduction of specific subdomains of endogenous lines. The intercept of a line – the limit of the function z when $\overline{y}$ tends to zero – is denoted by z^0.

[13] This interpretation derives from Sen (1983): "absolute deprivation in terms of a person's capabilities relates to relative deprivation in terms of commodities, incomes and resources". This interpretation can also be found in Atkinson and Bourguignon (2001).

[14] Slope Less than One is necessary in order for the line to admit an EO meeting Translation Monotonicity, defined below.

- *Piecewise-linear lines*:
  There exists $\overline{y}^k \geq 0$ and $\bar{s} \geq 0$ such that for all $\overline{y} \leq \overline{y}^k$, we have $z(\overline{y}) = z^0$ and for all $\overline{y} > \overline{y}^k$ we have $z(\overline{y}) = z^0 + \bar{s}(\overline{y} - \overline{y}^k)$.

- *Monotonic lines*:
  For all $\overline{y}, \overline{y}' > 0$ with $\overline{y} < \overline{y}'$, we have $z(\overline{y}) \leq z(\overline{y}')$ and there exists $g > 0$ with $g \geq z(g)$ such that $s(g) > 0$.

Finally, *linear lines* are piecewise-linear lines such that $\overline{y}^k = 0$, *relative lines* are linear lines such that $z^0 = 0$ and $\bar{s} > 0$ and *absolute lines* are linear lines such that $z^0 > 0$ and $\bar{s} = 0$.

Observe that weakly relative lines (Ravallion and Chen, 2011) are piecewise-linear lines. Both weakly relative lines and hybrid lines (Foster, 1998) are monotonic lines.

Equivalence orderings

I introduce here the concept of an equivalence ordering (EO). The set of accessible bundles is

$$X \coloneqq \{(y_i, \overline{y}) \in \mathbb{R}_+ \times \mathbb{R}_{++}\}.$$

These bundles are two-dimensional since the individual poverty of an agent depends on both her income and mean income in her society. Given an endogenous line z, the set of bundles at which an agent qualifies as poor is

$$X_p \coloneqq \left\{(y_i, \overline{y}) \in X \,\middle|\, y_i < z(\overline{y})\right\}.$$

An EO is a preference relation for an ethical observer that compares individual bundles. See Figure 4.4 for examples of EOs.

Definition 9 (Equivalence Ordering)**.**
An equivalence ordering $\succeq$ is a continuous ordering on X_p.[15]

The poverty line compares bundles in different societies. An EO extends this logic below the poverty line. The line is the frontier equivalence curve of the EO, defining the threshold below which an agent is deemed poor. Let $(y_i, \overline{y}) \succeq (y_i', \overline{y}')$ denote the judgment that agent i with income y_i when mean income is $\overline{y}$ has a weakly smaller individual poverty than with income y_i' when mean income is $\overline{y}'$. The symmetric and asymmetric parts of $\succeq$ are denoted $\sim$ and $\succ$ respectively.

The selection of an EO is a normative choice. This choice is arbitrary to some extent. As a consequence, some might be afraid that this approach leads to a significant increase in the arbitrariness of poverty judgments. This is not the case for two reasons. First, all major poverty measures implicitly define such an EO. Their EO is constrained by axioms with little ethical content such as *Scale Invariance*.[16]

[15] An ordering is a reflexive, transitive and complete binary relation.

[16] Another restriction implicitly constraining the EO is Translation Invariance, which requires the index not to be affected if the income of all agents is increased by the same amount as the poverty line.

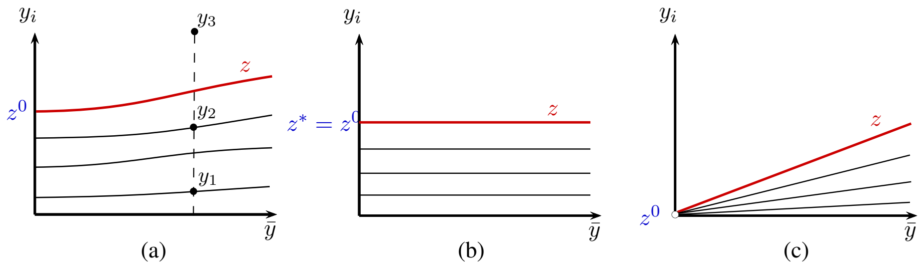

Figure 2.1: (a) Equivalence curves of a generic EO in $\mathcal{R}$ and an income distribution $(y_1, y_2, y_3) \in Y$. (b) EO implied by an absolute measure. (c) EO implied by a relative measure.

Second, the choice of an EO is guided by intuitive restrictions excluding exotic trade-offs between an agent's absolute and relative situation.

Given an endogenous line, the basic domain $\mathcal{R}$ of continuous EO is defined by three restrictions. Strict Monotonicity in Income requires that, at any mean income, more income leads to strictly smaller individual poverty.

EO restriction 1 (Strict Monotonicity in Income)**.**
For all $(y_i, \overline{y}) \in X_p$ *and* $a > 0$*, we have* $(y_i + a, \overline{y}) \succ (y_i, \overline{y})$.

The other two restrictions limit the importance of a poor agent's relative situation for her individual poverty. Translation Monotonicity requires that any poor agent is made weakly better-off by the equal distribution of an extra amount of income. It seems hard to conceive that the relative situation of a poor agent is made worse by such an equal distribution of income. As a result, the slopes of the EO's equivalence curves are never larger than one.

EO restriction 2 (Translation Monotonicity)**.**
For all $(y_i, \overline{y}) \in X_p$ *and* $a > 0$*, we have* $(y_i + a, \overline{y} + a) \succeq (y_i, \overline{y})$.

Finally, Minimal Absolute Concern requires that an agent with zero income is strictly poorer than another agent with non-zero income, regardless of the mean incomes in their respective societies.

EO restriction 3 (Minimal Absolute Concern)**.**
For all $(y_i, \overline{y}) \in X_p$ *with* $y_i > 0$ *and* $\overline{y}' > 0$*, we have* $(y_i, \overline{y}) \succ (0, \overline{y}')$.

Domain $\mathcal{R}$ is very wide and admits an infinity of different EOs below each endogenous line. This domain is flexible as it admits the implied EOs of standard poverty measures as special cases. This is for example the case of absolute and relative measures, as illustrated in Figure 4.4.b and 4.4.c.[17] As a result, additive indices presented in (2.1) are a special case of the family of additive indices based on endogenous lines that I derive below.

[17] The EO below a relative line in Figure 4.4.c lies inside the domain $\mathcal{R}$ since bundle $(0, 0)$ is excluded from the set X of bundles.

2.3.2 Characterization of an additive family

Any poverty index based on an absolute line and satisfying five basic axioms must have an additive mathematical expression (Foster and Shorrocks, 1991). I extend this standard result to indices based on endogenous lines. I present here the modified versions of the basic axioms used for this extension.

An EO captures the trade-offs at the individual level between the absolute and the relative situation. *Domination among Poor* requires poverty indices to respect the individual poverty comparisons encapsulated in an EO. It does so by imposing a monotonicity requirement in the space of individual poverty distributions, limited to poor agents. If the individual poverty of one poor agent decreases, while the individual poverties of all other agents do not increase, then poverty must decrease.

Social ordering axiom 13 (*Domination among Poor*).
There exists $\succeq \in \mathcal{R}$ *such that for all* $y, y' \in Y$ *with* $n(y) = n(y')$, *if* $(y'_i, \overline{y}') \succeq (y_i, \overline{y})$ *for all* $i \leq q(y')$, *then* $P(y) \geq P(y')$.
If in addition there is $j \leq q(y)$ *such that* $(y'_j, \overline{y}') \succ (y_j, \overline{y})$, *then* $P(y) > P(y')$.

Observe that *Domination among Poor* implies a weak version of *Focus*. Indeed, only the situation of poor agents is relevant for the index, but the incomes of non-poor agents can influence the index via the income standard.

Subgroup consistency is a standard axiom requiring that, if poverty decreases in a subgroup while it remains constant in the rest of the distribution, overall poverty must decline.[18] Sen (1992) questioned the desirability of this axiom by arguing that incomes in one subgroup may affect poverty in another subgroup. Foster and Sen (1997) recommend not to use this axiom when the index aims at capturing relative aspects of income poverty. I subscribe to this point of view. The issue becomes transparent once the channel through which one subgroup affects the other is modeled. In this framework, incomes in a subgroup impact mean income which affects other poor agents' individual poverty. If the line is absolute and the EO features flat equivalence curves (see Figure 4.4.b), relative income does not matter and subgroup consistency is compelling. If relative income does matter, then it is not always meaningful to extrapolate the judgments made on subgroups to the whole population. *Weak Subgroup Consistency* restricts such extrapolations to cases for which the incomes in a subgroup do not influence the individual poverty of agents in the other subgroup. These cases occur when the two subgroups of a population have the same mean income, implying that the subgroups have the same mean income as the total population. In such cases, poverty judgments made on subgroups are relevant for the total population.

Social ordering axiom 14 (*Weak Subgroup Consistency*).
For all $y^1, y^2, y^3, y^4 \in Y$ *such that* $n(y^1) = n(y^3)$, $n(y^2) = n(y^4)$, $\overline{y}^1 = \overline{y}^2$ *and* $\overline{y}^3 = \overline{y}^4$, *if* $P(y^1) > P(y^3)$ *and* $P(y^2) = P(y^4)$, *then* $P(y^1, y^2) > P(y^3, y^4)$.

[18] A formal definition of this axiom can be found in Foster and Shorrocks (1991).

The remaining three auxiliary axioms need no specific modification. *Symmetry* requires that agents identities do not matter. Working with sorted distributions is therefore without loss of generality.

Social ordering axiom 15 (*Symmetry*).
For all $y, y' \in Y$, *if* $y' = y \cdot \pi_{n(y)\times n(y)}$ *for some permutation matrix* $\pi_{n(y)\times n(y)}$, *then* $P(y) = P(y')$.

Continuity requires poverty indices to be continuous in incomes. This is important for empirical applications in order to avoid measurement errors having excessive impacts on poverty judgments.

Social ordering axiom 16 (*Continuity*).
For all $y \in Y$, P *is continuous in* y.

Replication Invariance permits comparing poverty in distributions of different population sizes. If a distribution is obtained by replicating another one several times, then the latter's poverty equals that of the original distribution.

Social ordering axiom 17 (*Replication Invariance*).
For all $y, y' \in Y$, *if* $n(y') = kn(y)$ *for some positive integer* k *and* $y' = (y, y, \ldots, y)$, *then* $P(y) = P(y')$.

Those five axioms allow us to derive an extension of the additive separability result of Foster and Shorrocks (1991). Its formal statement needs two definitions. First, a *numerical representation* is a continuous function representing an EO.

Definition 10 (Numerical Representation d).
The continuous function $d : X \to [0, 1]$ *is a numerical representation of* $\succeq \in \mathcal{R}$ *if*

- *for all* $(y_i, \overline{y}), (y'_i, \overline{y}') \in X_p$ *we have* $(y_i, \overline{y}) \succeq (y'_i, \overline{y}') \Leftrightarrow d(y_i, \overline{y}) \leq d(y'_i, \overline{y}')$,
- *for all* $(y_i, \overline{y}) \in X \backslash X_p$ *we have* $d(y_i, \overline{y}) = 0$.

A numerical representation differs from a utility representation of equivalence levels in two ways. First, it is constant for all equivalence levels above the poverty threshold. Second, below the poverty threshold, its value *decreases* when individual poverty decreases. The values returned by this function can be interpreted as individual poverty, i.e. the opposite of utility.

Next, I define additive poverty indices which aggregate agents' individual poverty by summing them.

Definition 11 (Additive Poverty Index).
P *is an additive poverty index if it is ordinally equivalent to another index* $\hat{P} : Y \to [0, 1]$ *such that for all* $y \in \mathbb{R}^N_+$

$$\hat{P}(y) := \frac{1}{n} \sum_{i=1}^{n} d(y_i, \overline{y}), \tag{2.4}$$

where d *is a numerical representation of an EO in* $\mathcal{R}$.

Theorem 1 characterizes the family of additive poverty indices based on endogenous lines. This is the first characterization of indices based on endogenous lines.

Theorem 1 (Characterization of additive poverty indices)**.**
Let P be a poverty index based on an endogenous poverty line. The following two statements are equivalent.

1. *P is an additive poverty index.*
2. *P satisfies Domination among Poor, Weak Subgroup Consistency, Symmetry, Continuity and Replication Invariance.*

Proof. It is easy to check that additive poverty indices satisfy these five axioms, so the proof that statement 1 implies statement 2 is hence omitted. The proof of the reverse implication is in Appendix 2.9.1. In a nutshell, the proof shows that the result on additive separability of Gorman (1968) applies. The crucial assumption to verify is that the index satisfies a separability property. After applying Theorem 1 in Gorman (1968), the remaining part of the proof is a modification of Foster and Shorrocks (1991). ■

The difference with the result of Foster and Shorrocks (1991) is that numerical representations of individual poverty depend now on two-dimensional bundles, made of own income and mean income. This new dependence on mean income vanishes if the line is absolute and the EO has only flat equivalence curves.

The family of additive indices is very broad. Choosing an index in that family requires selecting both an EO below the line and a numerical representation for this EO. I show in the next section that both normative choices can be deduced from largely shared intuitions. A new index emerges then as the focal additive index with good properties.

2.4 A new index with good properties

In this section, I first describe how to select an EO and its numerical representation from largely shared intuitions. Then, I present the index defined by these choices and show that this index is workable, it distinguishes absolutely poor agents from relatively poor agents and it is decomposable between the contributions of these two kinds of poor agents.

2.4.1 Selection of an equivalence ordering

The example given in Table 2.2 showed that endogenous measures satisfying *Scale Invariance* give no priority to subsistence over social participation. The EO of an endogenous measure satisfying *Scale Invariance* has its equivalence curves evolve as constant fractions of the income threshold. Geometrically, these equivalence curves are homothetic.

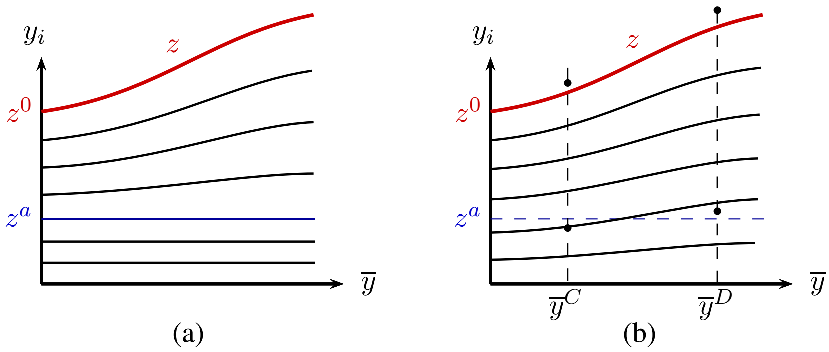

Figure 2.2: (a) Absolute-homothetic EO below the endogenous line z. (b) Homothetic EO below the endogenous line z and two income distributions y^C and y^D in Y. The homothetic EO deems the poor agent in distribution y^C to be less poor than the poor agent in distribution y^D, even if the latter earns income above the absolute threshold z^a.

An EO giving priority to subsistence can never consider that an agent whose income is below the subsistence threshold is less poor than another agent whose income is above, regardless of the standards of living in their respective societies. Accordingly, when comparing two agents with incomes below subsistence, the one with larger income can not be judged poorer than the other, independently of their respective relative situations. Only EOs having all their equivalence curves *flat* up to the subsistence threshold satisfy these intuitions.[19]

I define a subdomain of EOs based on a fourth restriction. Restriction Absolute-Homotheticity, illustrated in Figure 2.2.a, is defined from the absolute threshold z^a. This parameter can be interpreted as the subsistence threshold or alternatively as the threshold for absolute material deprivation. EOs satisfying Absolute-Homotheticity have all their equivalence curves flat up to z^a. This condition is formally expressed in part (i).

EO restriction 4 (Absolute-Homotheticity)**.**
There exists $z^a \geq 0$ such that for all $(y_i, \overline{y})$, $(y'_i, \overline{y}') \in X_p$:

(i) Priority to subsistence over social participation.
if $y_i = y'_i \leq z^a$ then $(y_i, \overline{y}) \sim (y'_i, \overline{y}')$,

(ii) Homothetic equivalence curves above the absolute threshold.
if $y_i, y'_i \geq z^a$ and $\frac{y_i - z^a}{z(\overline{y}) - z^a} = \frac{y'_i - z^a}{z(\overline{y}') - z^a}$, then $(y_i, \overline{y}) \sim (y'_i, \overline{y}')$.

(iii) Cost of social participation is never zero.
if $z^a > 0$, then $z^a < z^0$.

[19] I assume that the cost of subsistence – and hence the subsistence threshold – does not evovle with the standard of living.

Part (ii) of the restriction requires the equivalence curves above z^a to evolve as constant fractions of the distance between z^a and the income threshold. This simplifying assumption is a natural default option in the absence of reasons to deviate from it. In Section 2.5, I show that if equivalence curves above z^a deviate too much from homotheticity, then there exists no numerical representation of the EO with good properties.

The last part is technical and needed for some results, even though it has some normative content. Part (iii) requires the cost of social participation to be strictly positive, even in low-income societies. Ravallion (2012) defends this point by giving several examples of expenditures playing a social role in low-income countries such as festivals and celebrations.

Given that restriction Absolute-Homotheticity incorporates important intuitions, the selected EO should meet this restriction. Before turning to the selection of a numerical representation for an absolute-homothetic EO, I make two important remarks on this subdomain of EOs.

Given an endogenous line, the only parameter of the absolute-homothetic subdomain is the absolute threshold z^a. A particularly interesting feature of absolute-homothetic EOs is that they allow categorizing poor agents between those that are absolutely poor and those that are "only" relatively poor. An equivalent categorization can be found in Foster et al. (2013). This categorization implies that relatively poor agents are never absolutely poor, contrary to the categorization obtained when identifying the poor with two lines that cross (Atkinson and Bourguignon, 2001).

Let the *homothetic* EO be the absolute-homothetic EO for which $z^a = 0$. As illustrated in Figure 2.2.b, the homothetic EO does not give priority to subsistence over social participation. Absolute measures are based on the homothetic EO below an absolute line. Standard endogenous measures are based on the homothetic EO below an endogenous line. This shows that the absolute-homothetic domain of EOs generalizes the implied EOs of standard measures.

2.4.2 Selection of a numerical representation

Having selected an absolute-homothetic EO by fixing the absolute threshold, the only element of the additive index remaining unspecified is the EO's numerical representation. Many numerical representations should be discarded for their counter-intuitive judgments. I consider two properties that strongly constrain the set of acceptable numerical representations.

The first property is specific to poverty indices considering both the absolute and relative aspects of income. In such a framework, increasing the income of an agent entails a worse relative situation for the others. Poverty indices must balance those gains and losses without giving excessive importance to relative losses. *Monotonicity in Income* requires that decreasing the income of some *poor* agent never leads to an unambiguous poverty reduction.

Social ordering axiom 18 (*Monotonicity in Income*).
For all $y, y' \in Y$, if $y_i < y'_i < z(\overline{y}')$ and $y'_j = y_j$ for all $j \neq i$, then $P(y) \geq P(y')$.

When a poor agent's income increases, her individual poverty decreases as both her absolute and relative situation improve. On the other hand, mean income increases and this might increase the individual poverty of other poor agents, depending on the EO.[20] Moreover, the income threshold might increase and some agents who were non-poor might therefore become considered as poor. *Monotonicity in Income* requires that the positive impact of such an income increase is dominant. Observe that the larger the number of agents, the lower is the impact of such an income increase on mean income and hence on the individual poverty of others.

It is worth emphasizing that the Head-Count Ratio, when combined with an endogenous line, can conclude that destroying part of the income of a poor agent reduces poverty. The problem is illustrated in Table 2.3. The relative line z has its threshold equal to 50% of mean income. The distribution in society F is obtained from the distribution in society E by decreasing the income of poor agent 1. Nevertheless, the HC concludes there is more poverty in society E than in F.

Table 2.3: Index HC violates *Monotonicity in Income*.

	y_1	y_2	y_3	$z(\overline{y})$	$HC(y)$
Society E	2.5	3	12.9	3.1	$\frac{2}{3}$
Society F	2	3	12.9	2.9	$\frac{1}{3}$

The second property is a standard requirement that most poverty indices satisfy. *Transfer among Poor* requires that a Pigou-Dalton transfer taking place between two poor agents never unambiguously increases poverty.[21] This property is still very compelling when using mean income as an income standard since balanced transfers do not alter the mean. As a result, the individual poverty of agents not involved in the transfer is preserved.

Social ordering axiom 19 (*Transfer among Poor*)**.**
For all $y, y' \in Y$ and $\lambda > 0$, if $y_j - \lambda = y'_j > y'_k = y_k + \lambda$, $z(\overline{y}) > y_j$ and $y'_i = y_i$ for all $i \neq j, k$, then $P(y) \geq P(y')$.

I investigate which additive indices respect both properties. It is well-known that poverty indices satisfying *Transfer among Poor* are based on convex numerical representations. *Monotonicity in Income* is a new axiom in this context and I show below that it has a strong discriminative power.

A central result of this paper is that – when selecting an absolute-homothetic EO below a monotonic line – there is a unique numerical representation belonging to the *extended Foster-Greer-Thorbecke family* that satisfies both *Monotonicity in Income* and *Transfer among Poor*. Before formally stating this result in Theorem 2, I define this family of numerical representations.

[20]The individual poverty of absolutely poor agents is not affected as their equivalence curve is flat for absolute-homothetic EOs.

[21]A Pigou-Dalton transfer is a progressive balanced transfer preserving the relative ranks of the two agents involved in the transfer.

Beyond the subdomain of homothetic EOs, the graph of a numerical representation depends on the particular mean income at which it is drawn. As a result, the mathematical expression of a numerical representation depends on the reference mean income, denoted by $\overline{y}^r$, at which it is expressed. Defining a particular family of numerical representations requires introducing a function that specifies for each bundle the income yielding the same individual poverty at the reference mean income.[22]

Definition 12 (Equivalent Income Function at $\overline{y}^r$).
For any $\succeq \in \mathcal{R}$ and $\overline{y}^r > 0$, the equivalent income function $e^r : X \to [0, z(\overline{y}^r)]$ is the continuous function such that for all $(y_i, \overline{y}) \in X$: $(y_i, \overline{y}) \sim \big(e^r(y_i, \overline{y}), \overline{y}^r\big)$.

Given the restrictions on the domain $\mathcal{R}$ of EOs, the equivalent income function is well-defined.[23] Using the concept of equivalent income function, I propose an extension of the Foster-Greer-Thorbecke (FGT) family of numerical representations.

Definition 13 (Extended FGT Family).
For any given EO in $\mathcal{R}$, the numerical representation d belongs to the extended FGT family if there exist $\overline{y}^r \geq 0$ such that for all $(y_i, \overline{y}) \in X_p$:

$$d(y_i, \overline{y}) = \left(\frac{z(\overline{y}^r) - e^r(y_i, \overline{y})}{z(\overline{y}^r)}\right)^{\alpha} \qquad \text{with } \alpha \geq 0,$$

where e^r is the equivalent income function at $\overline{y}^r$.

The extended FGT family depends on two parameters: the reference mean income $\overline{y}^r$ at which d takes an exponential expression and the exponent α, interpreted as poverty aversion. For homothetic EOs, this family coincides with the standard FGT family presented in (2.2) since the mathematical expression of their numerical representation does not depend on the reference mean income.

In the extended FGT family, each value of poverty aversion defines a subfamily whose members are parameterized by the reference mean income. For example, the *PGR at $\overline{y}^r$* is the numerical representation that is linear ($\alpha = 1$) at mean income $\overline{y}^r$. The *PGR at the origin*, defined by $\overline{y}^r = 0$ and illustrated in Figure 2.3, plays a key role in the remainder of this paper.

Theorem 2 formalizes the central result showing that in the extended FGT family, only the PGR at the origin satisfies *Monotonicity in Income* and *Transfer among Poor*.

Theorem 2 (Characterization of PGR at the origin).
Let z be a monotonic line. Let P be an additive poverty index based on an absolute-homothetic EO below z with a numerical representation in the extended FGT family.

[22] As the case $\overline{y} = 0$ is ruled out from my domain, the definition of this function must be modified if $\overline{y}^r = 0$. The function e^0, the equivalent income function at $\overline{y} = 0$, is defined from $e^{\overline{y}'}$ with $\overline{y}' > 0$. Take any $(y_i, \overline{y}) \in X$, $e^0 : X \to [0, z^0]$ is the continuous function such that for all $\epsilon > 0$ there is $\delta > 0$ such that if $\overline{y}' < \delta$, then $|e^0(y_i, \overline{y}) - e^{\overline{y}'}(y_i, \overline{y})| < \epsilon$.

[23] The existence of an equivalent income function at *any* value of mean income is guaranteed for all EO's in our domain by restriction Minimal Absolute Concern. Furthermore, it is a function – it returns a unique value – since EOs meet restriction Strict Monotonicity in Income and since its domain of images is bounded above by the income threshold.

1. *P satisfies Monotonicity in Income only if:*

$$\alpha = 1.$$

2. *P satisfies Monotonicity in Income and Transfer among Poor if and only if:*

$$\alpha = 1 \qquad \textit{and} \qquad \overline{y}^r = 0,$$

that is, d is the PGR at the origin.

Proof. See in Appendix 2.9.3. The proof is based on Lemma 6, presented in Appendix 2.9.2, which gives a necessary condition and a sufficient condition for an index to satisfy *Monotonicity in Income*. ■

Theorem 2 shows that a unique member of the very rich extended FGT family satisfies both properties.

Claim 1 shows that *Monotonicity in Income* is responsible for the largest part of the result. First, among all values of poverty aversion, only the one associated to the PGR is acceptable. This characterization of the poverty aversion's value is due to the exponential mathematical form of the extended FGT family. For the case $\alpha < 1$, as the income of a poor agent tends to the income threshold, the priority granted to her over – say – an absolutely poor agent tends to infinity. Therefore, when the income of an absolutely poor agent increases, the individual poverty of a relatively poor agent close to the income threshold is negatively affected and the index concludes poverty has increased. The case $\alpha > 1$ is plagued with the reverse problem. As the income of a poor agent tends to the income threshold, her priority over other poor agents tends to zero. An increase in her income can be negatively judged by the index. Second, not all members of the PGR at $\overline{y}^r$ subfamily satisfy *Monotonicity in Income*. If the monotonic line is piecewise-linear, then the PGR at $\overline{y}^r$ satisfies the axiom if and only if $\overline{y}^r$ is below an upper-bound whose value depends on the parameters of the line and the absolute threshold.[24]

Claim 2 shows that *Transfer among Poor* further restricts the acceptable members of the PGR subfamily to a unique index. If the reference mean income is not $\overline{y}^r = 0$, then there exist mean incomes at which the numerical representation is concave, which violates *Transfer among Poor*. To see why, consider Figure 2.3. When drawn at the reference mean income, the graph of the PGR at $\overline{y}$ is linear, as shown in Figure 2.3.b for the PGR at the origin. When drawn at a larger mean income than the reference, its graph is piecewise-linear and convex because the income threshold is then larger than at the reference mean income, as shown in Figure 2.3.c for the PGR at the origin. If the reference value for mean income is not zero, then there exists values of mean income at which the income threshold is lower than at the refence mean income and the graph is piecewise-linear and concave.

[24]The proof for this claim can be found in Appendix 2.9.3. The intuition for the upper-bound goes as follows. The larger the reference mean income $\overline{y}^r$, the lower is the individual poverty gain made when bringing an agent with zero income to the absolute threshold. In other words, the larger $\overline{y}^r$, the lower the priority of absolutely poor agents over relatively poor agents. This priority tends to zero when $\overline{y}^r$ tends to infinity. In this sense, the upper-bound requires the index to guarantee a minimal priority to absolutely poor agents.

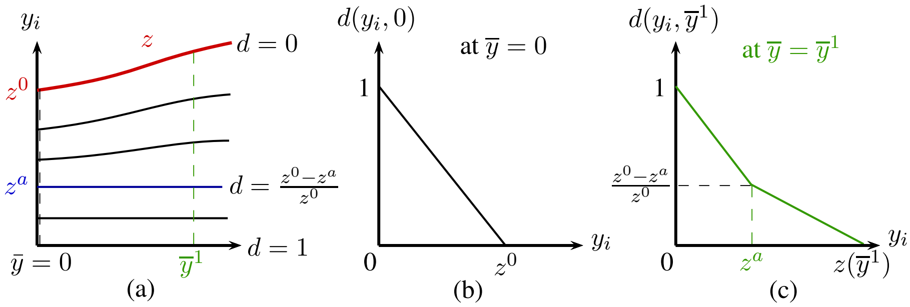

Figure 2.3: (a) Poverty Gap Ratio at the origin representing an absolute-homothetic EO (its values d are indicated at the end of three equivalence curves). (b) Graph of the PGR at the origin drawn at $\overline{y} = \overline{y}^r = 0$. (c) Graph of the PGR at the origin drawn at $\overline{y} = \overline{y}^1$.

The PGR at the origin representing an absolute-homothetic EO defines a new index of poverty. I present this new index in the coming subsection.

2.4.3 Presentation of the new index

Given an endogenous line z, how can the practitioner compute the index identified above?

The first step is to select an absolute-homothetic EO by fixing the absolute threshold z^a. The value for parameter z^a is selected to be either the subsistence threshold or a meaningful threshold for absolute material deprivation for the empirical question tackled by the practitioner. The choice of z^a defines an absolute-homothetic EO. The second step is mechanical and simply amounts to computing the mathematical expression of the index P. This mathematical expression, illustrated in Figure 2.3, is the PGR at the origin for the selected EO.

In practice, from distribution y with mean income $\overline{y}$, compute the censored distribution $\hat{y}$ by setting the income of all non-poor agents equal to $z(\overline{y})$. Compute then the *equivalent gap distribution* g^0 from the censored distribution $\hat{y}$. The equivalent gap of agent i is defined as:

$$g_i^0 := \frac{z^0 - e^0(\hat{y}_i, \overline{y})}{z^0}, \tag{2.5}$$

where $e^0(\hat{y}_i, \overline{y})$ is the equivalent income at $\overline{y} = 0$ given the selected EO, and z^0 is the intercept of the endogenous line. See Section 2.7 for an empirical illustration.

The new index is then simply the average equivalent gap:

$$P(y) := \frac{1}{n} \sum_{i=1}^{n} g_i^0.$$

Notice that the equivalent gap is different from the normalized gap presented in (2.3). The conceptual difference is that the former gives priority to subsistence over social participation by comparing individual situations using an absolute-homothetic

EO. The equivalent gap is only equal to the normalized gap in the special case in which the EO is homothetic ($z^a = 0$), thereby denying the existence of an absolute form of poverty.

Interesting features of the index

I argue in what follows that this new index is conceptually simple, it yields judgments in line with intuitions and it is decomposable between absolute and relative poverty.

The index is conceptually simple for two main reasons. First, the index makes a clear distinction between absolutely poor agents and relatively poor agents. The latter are never considered to be poorer than the former. Then, its expression is the average equivalent gap, interpretable as the average individual poverty in the population.

This additive index satisfies both *Monotonicity in Income* and *Transfer among Poor*. Furthermore, the index inherits the judgments of its absolute-homothetic EO. I emphasize that, when comparing poverty using this index:[25]

- An extra dollar has the same impact on global poverty when it is given to an absolutely poor agent in a low-income country as when it is given to an absolutely poor agent in a high-income country.

- An extra dollar has more impact on global poverty when it is given to a relatively poor agent in a low-income country than when it is given to a relatively poor agent in a high-income country. Even if bringing an agent from the subsistence threshold to the poverty threshold has the same impact on her individual poverty in both countries, it is more costly to do so in the high-income country.

- Growth, however unequally distributed, decreases the individual poverty of absolutely poor agents.[26]

- On the contrary, growth should not be too unequally distributed in order for the individual poverty of relatively poor agents to decrease.

A corollary of the last two bullet points is that this index concludes that growth, if strong enough, eventually eradicates absolute poverty but not necessarly relative poverty. Whether the latter form of poverty is eventually eradicated depends on the distributive aspects of growth.

Finally, this additive index is decomposable between the absolute and relative aspects of income poverty. Absolutely poor agents have income below the absolute threshold z^a, whereas relatively poor have income between the absolute threshold and the poverty threshold in their society. The numerical representation attributes

[25] Remember that from a theoretical perspective, comparing two different countries or the same country at different points in time is equivalent.

[26] This judgment resonates with the ideas of Sen (1983): "If there is starvation and hunger, then - no matter what the relative picture looks like - there clearly is poverty. In this sense, the relative picture – if relevant – has to take a back seat..."

an individual poverty equal to zero for non-poor agents and equal to one for agents with zero income. The key parameter $\frac{z^a}{z^0}$ measures which fraction of the zero-one range is attributed to absolute poverty. This fraction corresponds to the evolution of the individual poverty of an agent, from 1 to $1 - \frac{z^a}{z^0}$, when bringing her income from zero to the absolute threshold. The complement of this fraction is attributed to relative poverty. Hence, the individual poverty $d(y_i, \overline{y})$ of an *absolutely poor* agent can be decomposed between its absolute contribution $d^{aa}(y_i, \overline{y})$ and its relative contribution $d^{ar}(y_i, \overline{y})$,

$$d(y_i, \overline{y}) = d^{aa}(y_i, \overline{y}) + d^{ar}(y_i, \overline{y}),$$

where

$$d^{aa}(y_i, \overline{y}) := \frac{z^a - y_i}{z^0},$$
$$d^{ar}(y_i, \overline{y}) := \frac{z^0 - z^a}{z^0}.$$

The individual poverty of a *relatively poor* agent is directly equal to its relative contribution $d^r(y_i, \overline{y})$. Let q^a be the number of absolutely poor agents in y. By definition, the number of relatively poor agents equals $q - q^a$. The index can be decomposed in the following way:

$$P(y) = P^a(y) + P^r(y), \tag{2.6}$$

where

$$P^a(y) := \frac{1}{n}\left(\underbrace{\sum_{i=1}^{q^a} d^{aa}(y_i, \overline{y})}_{1} + \underbrace{\sum_{i=1}^{q^a} d^{ar}(y_i, \overline{y})}_{2}\right), \tag{2.7}$$

$$P^r(y) := \frac{1}{n}\left(\underbrace{\sum_{i=q^a+1}^{q} d^r(y_i, \overline{y})}_{3}\right). \tag{2.8}$$

Index P is hence decomposable between the contribution P^a of absolutely poor agents and the contribution P^r of relatively poor agents. The contribution of absolutely poor agents can be further decomposed. Term 1 in (2.7) measures the absolute contribution due to their individual poverty, coming from earning less than the subsistence threshold. Term 1 is ordinally equivalent to the PGR based on an absolute line whose threshold is the absolute threshold. Term 2 measures the relative contribution due to the individual poverty of absolutely poor agents. Term 3 in (2.8) accounts for the individual poverty of relatively poor agents, and therefore measures the relative poverty of relatively poor agents. Terms 2 and 3 together measure the total relative poverty in the population.

A last remark relates to the key parameter $\frac{z^a}{z^0}$. Given a particular poverty line, the domain of absolute-homothetic EOs has the absolute threshold as the unique parameter. Therefore, the index proposed is technically a family of indices, parameterized by $\frac{z^a}{z^0}$. If this fraction tends to one, the index tends to consider absolute and relative poverty in lexicographic order. In this case, any two income distributions are first compared based on the absolute poverty of absolutely poor agents, using term 1 in (2.7). If the comparison is non-conclusive, then relative poverty enters the picture. On the other hand, if $\frac{z^a}{z^0}$ tends to zero, there exist no absolutely poor agents and the index becomes the standard PGR based on the endogenous line. These two limit positions are rather extreme and the value of this fraction should hence not deviate too much from one half. Most importantly, the parameter z^a should be a meaningful absolute threshold for the question tackled.

Given its interesting features, this index is a good candidate for comparing poverty between societies having different standards of living. I conduct an empirical application using this index in Section 2.7. In the next two sections, I investigate the robustness of Theorems 1 and 2 to several assumptions.

2.5 Robustness with mean income as the income standard

I show in this section that, when mean income is the income standard, any other index satisfying *Monotonicity in Income* and *Transfer among Poor* should be "close" to the index presented above.

2.5.1 Outside the extended FGT family

The very sharp conclusions of Theorem 2 are valid for numerical representations in the extended FGT family. I investigate in this subsection the robustness of these conclusions outside that family. I show by means of an example that, for other families, the discriminating power of *Monotonicity in Income* is less strong but the PGR at the origin still emerges as the focal numerical representation. Any other numerical representation satisfying *Monotonicity in Income* and *Transfer among Poor* must be close to the PGR at the origin.

For simplicity, the poverty line is linear and the EO is homothetic. Given these assumptions, I define the quadratic family of numerical representations. This family has no particular ethical appeal but is useful to illustrate the trade-off emerging from *Monotonicity in Income*.

Definition 14 (Quadratic Family)**.**
For any homothetic EO, the numerical representation d belongs to the quadratic family if for all $(y_i, \overline{y}) \in X_p$:

$$d(y_i, \overline{y}) = \left(1 - \frac{y_i}{z(\overline{y})}\right) + \alpha\left(\left(\frac{y_i}{z(\overline{y})}\right)^2 - \frac{y_i}{z(\overline{y})}\right) \qquad \textit{with } \alpha \in [-1, 1].$$

The quadratic family admits a unique parameter α interpreted as poverty aversion. The case $\alpha = 0$ corresponds to the standard PGR. Quadratic poverty indices satisfy *Domination among Poor* only when α belongs to $[-1, 1]$, a range which allows

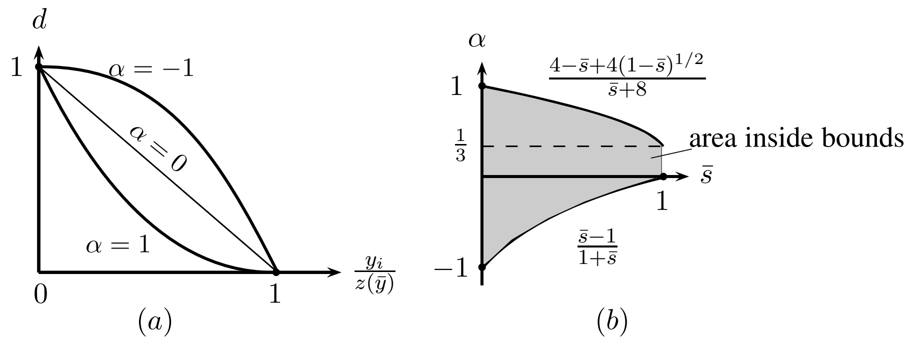

Figure 2.4: (a) Numerical representations in the quadratic family for three different values of the poverty aversion parameter α. (b) The upper and lower bounds on the poverty aversion parameter α evolve monotonically as a function of the slope $\bar{s}$ of the line.

for much less variety of judgments around the PGR than the extended FGT family. The restrictions on α under which *Monotonicity in Income* is satisfied are stated in Theorem 3 and illustrated in Figure 2.4.b. The coefficient of poverty aversion is bounded above and below and those bounds depend monotonically on the poverty line's slope.

Theorem 3 (Bounds on poverty aversion around PGR)**.**
Let z be a linear poverty line with slope $\bar{s}$. Let P be an additive poverty index based on a homothetic EO below z with a numerical representation in the quadratic family. P satisfies Monotonicity in Income if and only if:

$$\frac{(\bar{s}-1)}{(1+\bar{s})} \leq \alpha \leq \frac{4-\bar{s}+4(1-\bar{s})^{\frac{1}{2}}}{(\bar{s}+8)} \tag{2.9}$$

Proof. See in Appendix 2.9.4. ■

The steeper the slope, the narrower is the range of acceptable values for poverty aversion around the case $\alpha = 0$, corresponding to the PGR. There is no collapse towards the PGR when the slope is equal to one. There exist indices exhibiting a – slightly – higher poverty aversion than the PGR that respect *Monotonicity in Income* and *Transfer among Poor*.

Theorem 3 is obtained for homothetic EOs and linear lines. Defining the *extended* quadratic family using equivalent income functions, a similar bound result can be derived for any absolute-homothetic EO below a monotonic line. The PGR at the origin is not the only index satisfying the two properties. Outside the extended FGT family, there are acceptable indices with larger poverty aversion. However, this bound result shows that the numerical representation of alternative indices should not be too far from the PGR. The steeper the poverty line, the closer these indices are to the PGR at the origin. In this sense, the PGR at the origin is focal.

So far, the EO has been assumed absolute-homothetic. As argued in sections 2.2 and 2.4, there are good ethical reasons for flat equivalence curves below the absolute

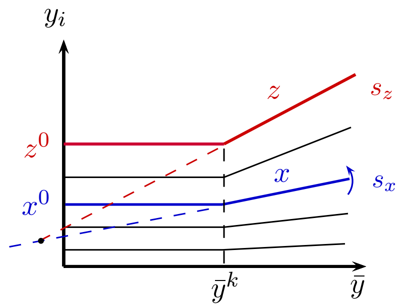

Figure 2.5: Homothetic-homothetic EO based on a piecewise-linear line.

threshold. Nevertheless, the homotheticity of equivalence curves above the absolute threshold has just been presented as a convenient assumption. A natural question to ask is whether choosing an EO from a different domain allows for a wider set of indices satisfying both *Monotonicity in Income* and *Transfer among Poor*. As shown in the next subsection, the mere existence of such indices is not guaranteed, even for EOs that "almost" belong to the absolute-homothetic domain. What is more, the PGR at the origin is still the focal index with good properties.

2.5.2 Beyond absolute-homothetic orderings

This section provides an additional reason to rely on absolute-homothetic EOs. Some additive indices based on absolute-homothetic EOs satisfy both *Monotonicity in Income* and *Transfer among Poor*.[27] There are no such indices if the EO departs too much from being absolute-homothetic.

For simplicity, the poverty line is piecewise-linear. For such lines, I define a domain extending the absolute-homothetic domain. The homothetic-homothetic domain $\mathcal{R}^{HH}$, illustrated in Figure 2.5, is defined from the general domain $\mathcal{R}$ by the additional restriction Homothetic-Homothetic Piecewise-Linear.

EO restriction 5 (Homothetic-Homothetic Piecewise-Linear)**.**
There exist two piecewise-linear curves x and z defined by:

$$x(\overline{y}) = \begin{cases} x_0 & \text{if } \overline{y} \leq \overline{y}^k, \\ x_0 + s_x(\overline{y} - \overline{y}^k) & \text{else}, \end{cases} \qquad z(\overline{y}) = \begin{cases} z_0 & \text{if } \overline{y} \leq \overline{y}^k, \\ z_0 + s_z(\overline{y} - \overline{y}^k) & \text{else}, \end{cases}$$

with $0 \leq s_x \leq s_z$, $0 < x^0 < z^0$ *and* $\overline{y}^k \geq z_0$ *such that for all* $(y_i, \overline{y})$, $(y'_i, \overline{y}') \in X_p$*:*

(i) Homothetic equivalence curves below x.
if $y_i < x(\overline{y})$ *and* $\frac{y_i}{x(\overline{y})} = \frac{y'_i}{x(\overline{y}')}$, *then* $(y_i, \overline{y}) \sim (y'_i, \overline{y}')$.

(ii) Homothetic equivalence curves between x and z.
if $y_i \geq x(\overline{y})$ *and* $\frac{y_i - x(\overline{y})}{z(\overline{y}) - x(\overline{y})} = \frac{y'_i - x(\overline{y}')}{z(\overline{y}') - x(\overline{y}')}$, *then* $(y_i, \overline{y}) \sim (y'_i, \overline{y}')$.

[27] I am grateful to Martin Ravallion for having pointed out that the existence of additive indices respecting *Monotonicity in Income* is not guaranteed for all EO in $\mathcal{R}$.

For a given poverty line z as defined in restriction 5 and $x^* < z^0$, the subdomain of EOs $\mathcal{R}^{HH}(z, x^*)$ is parameterized by the slope s_x of the intermediate line x defined in restriction Homothetic-Homothetic Piecewise-Linear:

$$\mathcal{R}^{HH}(z, x^*) := \left\{ \succeq \ \in \mathcal{R}^{HH} \,\middle|\, \succeq \text{ is below line } z \text{ and } x^0 = x^* \right\}.$$

Let $\succeq_{s_x}$ denote a generic element in $\mathcal{R}^{HH}(z, z^a)$. The case $s_x = 0$ corresponds to the absolute-homothetic EO at the origin of the subdomain $\mathcal{R}^{HH}(z, x^*)$. The case $s_x = s_x^h := \frac{x^0}{z^0} s_z$ corresponds to the homothetic EO below the poverty line. The case $s_x = s_z$ is the limit since larger intermediate slopes entail that both lines cross.

Theorem 4 consists of two claims. Claim 1 says there is a range of values for the slope parameter s_x, centered on the value $s_x = s_x^h$ making the EO homothetic, outside which no additive poverty index satisfies both properties. Claim 2 says that if the slope parameter s_x is smaller than s_x^h and if the numerical representation belongs to the extended FGT family, then the two properties force the numerical representation to be the PGR at the origin.

Theorem 4 (Non absolute-homothetic EOs and PGR at the origin)**.**
Let z be a piecewise-linear poverty line with $\overline{y}^k \geq z^0$ and $\bar{s} > 0$. Let $x^* > 0$ be such that $x^* < z^0$. Let $\succeq_{s_x}$ be an EO belonging to the subdomain $\mathcal{R}^{HH}(z, x^*)$.

1. *There exists an additive index P based on $\succeq_{s_x}$ satisfying both Monotonicity in Income and Transfer among Poor if and only if*
   - *either $s_x = 0$,*
   - *or for some $\underline{s}_x$ and $\overline{s}_x$ with $\underline{s}_x < s_x^h < \overline{s}_x < s_z$ we have:*[28]

   $$\underline{s}_x \leq s_x \leq \overline{s}_x.$$

2. *Assume $s_x \in [\underline{s}_x, s_x^h]$ with $s_x \geq 0$. Let P be an additive index based on $\succeq_{s_x}$ with a numerical representation belonging to the extended FGT family. The two following statements are equivalent:*
   - *P satisfies both Monotonicity in Income and Transfer among Poor.*
   - *The numerical representation of P is the PGR at the origin.*

Proof. See in Appendix 2.9.5. ■

The domain $\mathcal{R}^{HH}$ of homothetic-homothetic EOs is defined without flat equivalence curves below an absolute threshold. Extending the definition of the domain $\mathcal{R}^{HH}$ to a domain $\mathcal{R}^{AHH}$ of *absolute*-homothetic-homothetic EOs with $z^a < x^0$ is straightforward. The last result can then be extended to EOs in $\mathcal{R}^{AHH}$. This

[28]The expressions for $\underline{s}_x$ and $\overline{s}_x$ are respectively:

$$\underline{s}_x := s_z - \frac{z^0 - x^0}{x^0} \qquad \text{and} \qquad \overline{s}_x := \frac{1}{2}\left(\left(\left(\frac{x^0}{z^0 - x^0}\right)^2 + 4 s_z \frac{x^0}{z^0 - x^0}\right)^{0.5} - \frac{x^0}{z^0 - x^0}\right).$$

extended result implies for the definition of absolute-homothetic EOs that have homothetic curves above the subsistence threshold is not just a convenient assumption but rather a precondition for the existence of indices satisfying both *Monotonicity in Income* and *Transfer among Poor*. Other EOs admitting poverty indices with these properties are not too far from being absolute-homothetic, as shown by the acceptable range around the homothetic value of the slope parameter given in Claim 1 of Theorem 4.

I have made the claim that *Monotonicity in Income* cannot be satisfied by an index based on an EO far from satisfying restriction Absolute-Homotheticity. This raises the question of the relationship that poverty axioms and EO restrictions have. What is their relative status in the case that an incompatibility arises? In my view, they have an equal status in the sense that they both constrain the set of acceptable indices. The difference is the channel through which they constrain them. Axioms constrain the comparison of distributions whereas EO restrictions constrain the comparison of individual bundles. When an incompatibility arise between a set of axioms and EO restrictions, one must arbitrate between them on the basis of their respective normative merits. I see no reason to systematically give priority to one type of "index constraint" over the other.

The message of this section is that the PGR at the origin is the focal numerical representation satisfying both *Monotonicity in Income* and *Transfer among Poor*. This conclusion is derived when considering that mean income is the relevant reference statistic for standards of living. In the next section, I argue why mean income is a good income standard for poverty measurement. Moreover, I show that the index proposed is still very relevant when using other income standards.

2.6 Income standards other than the mean

I discuss in this section the choice of the income standard to which the poverty line is sensitive. I argue that median income is not a good income standard for poverty measurement and that other income standards are preferable, such as the mean or a lower partial mean. Finally, I study the robustness of Theorems 1 and 2 when using poverty lines sensitive to income standards different than mean income.

An income standard is a reference statistic gauging the size of an income distribution. Let $f : \mathbb{R}_+^N \to \mathbb{R}_+$ denote an income standard. As for poverty indices, income standards can be derived from the properties defining its concept. The two properties more specific to income standards are *Normalization* and *Linear homogeneity*. Normalization requires that if all incomes in a distribution are equal, then the income standard is also equal to the common income. Linear homogeneity requires that if all incomes in a distribution are multiplied by a common factor, the value taken by the income standard is also multiplied by the same factor.

Besides the mean, there exist four types of income standards that are in common use (Foster et al., 2013). These four types are quantiles (e.g. the median), generalized means (e.g. the geometric mean), partial means (e.g. mean among the 99% least rich individuals) and the Sen mean. All four types of income standards are presented and discussed in this section.

The choice of income standard is important because it defines the channel through which the income of other agents affect individual poverty. More specifically, it defines the distributional changes altering the income threshold and, hence, the individual poverty of poor agents. Poverty judgments hence depend on the income standard used.

I defend the use of mean-sensitive lines even if median-sensitive lines are often used in practice. For example, the AROP measure of the European Commission uses a median-sensitive line.[29] Both statistics have different advantages and flaws.

The main advantage of the mean corresponds to a major flaw of the median. A poverty measure aims at evaluating the impact that economic policies have on the worse-off individuals. If some policies impact growth, many policies have only redistributive consequences. I find highly counter-intuitive that policies whose unique impact are regressive transfers from the middle class to the rich are deemed to be poverty reducing. Axiom *Transfer among Non-Poor* requires the index not to be affected by redistributions among non-poor individuals. An index satisfying this axiom is therefore immune to "redistributive manipulations". *Transfer among Non-Poor* is a weakening of *Focus*.

Social ordering axiom 20 (*Transfer among Non-Poor*)**.**
For all $\delta > 0$ *and all* $y, y' \in Y$ *with* $n(y') = n(y)$*, if* $y_j - \delta = y'_j > y'_k = y_k + \delta$*,* $k > q(y)$ *and* $y'_i = y_i$ *for all* $i \neq j, k$*, then* $P(y) = P(y')$*.*

For mean-sensitive indices, *Transfer among Non-Poor* is implied by *Domination among Poor*. In contrast, de Mesnard (2007) has shown that median-sensitive indices behave very counter-intuitively when income distributions experience an increase in inequality. The issue does not only show up in theory, it is particularly problematic in a World in which intra-country inequalities are on the rise (Bourguignon, 2013). An illustration of such behavior took place in New-Zealand between 1981 and 1992. According to Easton (2002), the implementation of policies inducing regressive transfers led to a decrease in the income of the bottom 80 % of households. Nevertheless, the median-sensitive HC dropped due to the large decline in median income and some institutions used these figures to argue the regressive policies were a success.

The main drawback of mean-sensitive indices is that the mean is affected by "outliers". What if a policy incentivize a very rich individual – say Bill Gates – to immigrate to the country, or simply allows an individual to flourish and become very rich? This could be good news but some fear that a mean-sensitive index systematically concludes otherwise. In theory, this need not necessarily be the case. Indeed, the conclusion depends on the redistributive system of the country. If the presence of very rich benefits to the poor – say via the country's tax-and-transfers system – then the value returned by a mean-sensitive index can decrease. In this sense, a mean-sensitive index can be used to judge a country's institutions. In

[29] The *At Risk of Poverty* measure is the Head-Count Ratio based on a relative line whose threshold is 60 % of the median income.

practice, however, the income standard is not always computed from administrative data but often from random samples. The median is known to be more robust than the mean in random samples (Cowell and Victoria-Feser, 1994). Median-sensitive lines have hence a less volatile income threshold.

For those who judge that the lower robustness of the mean is a more serious issue in practice than the "manipulability" of the median, mean income among the 99% least rich individuals can offer a good compromise. This partial mean is much less affected by outliers than the mean. The downside of such a partial mean is that regressive redistributions among non-poor individuals benefiting the 1% richest individuals affect it.

In the remainder of this section, I study the robustness of the results to the use of different income standards. The median is first investigated before turning to other income standards.

2.6.1 Median income

The median is a particular quantile. Let $x \in [0, 100]$ be a percentile. The quantile income at the x^{th} percentile in distribution y is the income level y_x such that x percent of individuals earn more than y_x and $1 - x$ individuals earn less. Quantile incomes are crude as they only provide information about a specific point of the distribution.

Median income, corresponding to the case $x = 50$, is the income standard such that half of the population earn more and half of the population earn less. Formally, the definition of median income is slightly different for distributions with even or odd number of dimensions. Median income y_m is defined to be the income of agent m where $m : \mathbb{N}_+ \to \mathbb{N}_+$ is defined by:

$$m := \begin{cases} \frac{n+1}{2} & \text{if n is odd,} \\ \\ \frac{n}{2} & \quad\text{otherwise.} \end{cases}$$

Changing the income standard requires modifying several definitions. I present here only the major non-straightforward modifications. For a given median-sensitive poverty line z, the results depend on the domain of income distributions considered. Let Y^r be the domain of distributions containing a strict minority of poor agents and let Y^p be its complement:

$$\begin{aligned} Y^r &:= \{y \in \mathbb{R}^N_+ |\ z(y_m) \leq y_m\}, \\ Y^p &:= \{y \in \mathbb{R}^N_+ |\ 0 < y_m < z(y_m)\}. \end{aligned}$$

Given the median-sensitive line z, let y^*_m be the lowest value of median income for which $y \in Y^r$, implicitly defined by $z(y^*_m) = y^*_m$. This is the limit value for median income above which a distribution belongs to Y^r. Let $Y := Y^p \cup Y^r = \{y \in \mathbb{R}^n_+ |\ y_m > 0\}$ be the general domain of distributions. Poverty indices are based on an equivalence ordering $\succeq^m$ ranking the set of poor bundles

$$X_p := \{(y_i, y_m) \in X |\ z(y_m) > y_i\},$$

where $X := \mathbb{R}_+ \times \mathbb{R}_{++}$. Poverty axioms as well as restrictions to endogenous lines and EOs in the general domain $\mathcal{R}_m$ are easily modified. Such modifications allow characterizing additive poverty indices with median-sensitive lines for the domain of income distributions containing a strict minority of poor agents.

Theorem 5 (Characterization of median-sensitive additive poverty indices).
Let $P : Y \to \mathbb{R}$ be a poverty index based on a median-sensitive poverty line. Statement 2 implies statement 1.

1. *On Y^r, P is ordinally equivalent to an index $P' : Y \to [0, 1]$ defined by*

$$P'(y) = \frac{1}{n}\sum_{i=1}^{n} d(y_i, y_m), \tag{2.10}$$

   where d is a numerical representation of an EO in $\mathcal{R}_m$.

2. *P satisfies the modified versions of Domination among Poor, Weak Subgroup Consistency, Symmetry, Continuity and Replication Invariance.*

Proof. See in Appendix 2.9.6. ■

With median-sensitive lines, the characterization of additive poverty indices is only valid on Y^r. On the general domain Y, additive indices satisfy the five axioms, but there might be other indices to do so.

The consequences of the modified version of *Monotonicity in Income* are different than for mean-sensitive poverty lines. This axiom constrains the domain of median-sensitive lines for which there exists additive indices respecting it. Additive indices respect *Monotonicity in Income* when their poverty line is flat for all values of median income below y_m^*.

Theorem 6 (Flat median-sensitive lines for low median incomes).
Let z be a monotonic median-sensitive poverty line with $z^0 > 0$. Let $P : Y \to [0, 1]$ be an additive poverty index based on an absolute-homothetic EO below z. The following two statements are equivalent.

1. *P satisfies Monotonicity in Income.*

2. *z is flat for all $y_m < y_m^*$.*

Proof. See in Appendix 2.9.7. ■

For all distributions in Y^r, median income is above the poverty threshold and hence the incomes of poor agents do not affect the reference statistic. The modified version of *Monotonicity in Income* puts no extra constraint on additive indices as this axiom is implied by *Domination among Poor*. For distributions in Y^p, the median income is below the threshold. If the median income increases by an amount not sufficient for the median agent to change, the reference statistic changes by the same amount, irrespective of the number of agents. This drastic impact drives the result.

In contrast, mean income changes only by a fraction $\frac{1}{n}$ of the amount gained by a poor agent.

I argued above that median-sensitive lines lead to counter-intuitive judgments about the impacts of regressive redistributive policies. If their flaws are not judged serious enough for switching to mean-sensitive lines, this research still provides good reasons for adopting the index proposed in Section 2.4. Indeed, contrary to the HC or the PGR, this index gives priority to subsistence over social participation.

2.6.2 Other income standards

Besides the mean and the median, other income standards can be used as reference statistic. I discuss the robustness of the results in each case.

Partial means

Partial means return mean income for a subset of the distribution. Two types are in common use: lower partial means and upper partial means. As for quantiles, they are attached to a percentile $x \in [0, 100]$. In distribution y, the lower partial mean below x, denoted $f^{\ell pm}(y, x)$, is the mean income among the bottom x percent of income earners. On the contrary, the upper partial mean above x is the mean income among the top $100 - x$ percent of income earners.

I only consider lower partial means because they better capture the evolution of the cost of social participation for poor individuals. Furthermore, since lower partial means are not affected by outliers, lines sensitive to these income standards offer a good compromise between the issues attached to mean and median-sensitive lines, respectively.

The additive representation result holds for the lower partial mean below x if the set of distributions considered only contains distributions for which the percentage of poor individuals is less than x:

$$Y^{\ell pm} := \{y \in \mathbb{R}^N_+ | \ z\big(f^{\ell pm}(y)\big) \leq y_{\frac{x}{100}n}\},$$

where $y_{\frac{x}{100}n}$ denotes the income of the agent whose index i is the largest natural number less than or equal to $\frac{x}{100}n$.

The characterization of the PGR at the origin as the only numerical representation inside the FGT family satisfying modified versions of *Monotonicity in Income* and *Transfer among Poor* holds when using lower partial means. Formal statements and proofs of these two claims may be found in Appendix 2.9.8.

Generalized means

Generalized means form a class of income standards putting more emphasizis on the bottom or on the top of the income distribution, depending on the value taken by its unique parameter $\beta \in (-\infty, +\infty)$. Generalized means, denoted $f^{gm}(y, \beta)$, are defined in the following way (Atkinson, 1970):

$$f^{gm}(y,\beta) := \begin{cases} \left(\frac{y_1^\beta + \cdots + y_n^\beta}{n}\right)^{\frac{1}{\beta}} & \text{if } \beta \neq 0, \\ (y_1 \times \cdots \times y_n)^{\frac{1}{n}} & \text{if } \beta = 0. \end{cases}$$

If $\beta < 1$, then the bottom of the distribution is emphasized, if $\beta > 1$, then the top of the distribution is emphasized. The most popular members of this class are the arithmetic mean ($\beta = 1$), the geometric mean ($\beta = 0$) and the harmonic mean ($\beta = -1$).

I only consider generalized means with $\beta < 1$ as they better capture the evolution of the cost of social participation for poor individuals.[30] These generalized means are not well-suited income standards for my purpose. Any EO respecting modified versions of the basic restrictions defining $\mathcal{R}$ must have all its equivalence curve flat. Such an EO cannot account for the impact that the relative situation has on individual poverty.

Theorem 7 (Non-flat EO violates Translation Monotonicity)**.**
Let f^{gm} be an income standard in the generalized mean family with $\beta < 1$. If $\succeq$ is an EO respecting the modified version of Strict Monotonicity in Income and $\succeq$ is non-flat, then $\succeq$ violates the modified version of Translation Monotonicity.[31]

Proof. See in Appendix 2.9.9. ∎

This result is a consequence of the exponential expression of this income standard. A small increment given to an agent whose income is close to zero has a disproportionate impact on the value taken by the generalized mean. Then, the small increment received by another poor agent whose bundle is on a non-flat indiffence curve cannot compensate for this disproportionate increase in income standard. The non-flat EO considers that the equal distribution of the additional resource made this other agent poorer, which violates Translation Monotonicity.

Notice that Translation Monotonicity is imposed as a restriction on the EO rather than as a poverty axiom. Imposing that equal increments reduces poverty as an axiom would be less strong. When *Domination among Poor* is imposed, the EO restriction Translation Monotonicity implies this associated axiom. Nevertheless, this EO restriction appears as a minimal limitation at the individual level to the importance of relative aspects of income. Therefore, I conclude from Theorem 7 that generalized means should not serve as income standards for poverty measurement.

---

[30] As generalized means satisfy a separability property, the additive representation result should hold if the set of distributions considered only contains distributions for which at least one individual is non-poor.

[31] An EO is *non-flat* if there exists $(y_i, f^{gm}(y)) \in X_p$ such that $s(y_i, f^{gm}(y)) > 0$, i.e. the slope in $(y_i, f^{gm}(y))$ is strictly positive.

Sen mean

The Sen mean, denoted f^{sm}, is interpreted as the expected minimum value among two income draws with replacement in the distribution.

$$\begin{aligned} f^{sm}(y) &:= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} min\{y_i, y_j\} \\ &= \frac{1}{n^2}\left((2n-1)y_1 + (2n-3)y_2 + \cdots + 3y_{n-1} + y_n\right). \end{aligned}$$

By definition, this income standard cannot be larger than the mean income. The Sen mean places more emphasis on incomes at the bottom of the distribution.

Since the Sen mean does not satisfy a separability property, the equivalent of the additive representation theorem does not hold. In particular, additive indices do not satisfy the modified version of *Weak Subgroup Consistency*.

Unlike for generalized means, the modified version of Translation Monotonicity only requires the equivalence curves of the EO to have slopes no larger than one.[32] Using the Sen mean as the income standard therefore allows us to account for the impact that relative income has on individual poverty. Neveretheless, the exact implications of the modified version of *Monotonicity in Income* for Sen-mean-sensitive indices is still an open question.

Moving average of an income standard

I discuss in this subsection an important point valid for any choice of income standard. Endogenous measures are regularly criticized for the counter-intuitive judgments they sometimes provide when a distribution is affected by a negative shock. Think of a transient economic crisis. Assume that even if all incomes decrease, the crisis has a smaller effect on the incomes at the bottom of the distribution than at the top. Endogenous measures can conclude that poverty has decreased, a highly debatable judgment. This problem is of course coming from the endogeneity of the line.

Based on such examples, some argue against the use of endogenous measures. Instead, they suggest using absolute measures whose absolute line is unchanged over many years and then updated to account for changes in the standard of living. After having changed the line, comparisons across the two periods – the periods before and after the update – are typically made using the new absolute line. It should be clear that this approach does not account for social participation effects (illustrated in Table 2.1). As a result, growth is deemed poverty reducing, regardless of how unequally distributed its gains are.

Another point of view on the counter-intuitive judgments made by endogenous measures is that the income standard is not appropriate. In practice, endogenous lines have their income threshold updapted each time the poverty measure is recomputed, typically every year. The cost of social participation has some inertia and

[32] For all $y \in \mathbb{R}^N_+$, given the mathematical expression of the Sen mean, we have $(\nabla f^{sm}(y) \cdot \mathbf{1_n}) = \mathbf{1}$.

does not react as quickly. At the beginning of a crisis, poor individuals have a lower income and face almost the same costs of social participation. It takes some time before people adapt their social standards and expectations. A solution would then be to introduce some inertia in the income standard. This can be done by letting the poverty line evolve with a moving average of the values taken by the income standard over several years.[33]

This section has discussed the choice of an income standard. I emphasized that median-sensitive lines lead to counter-intuitive judgments when intra-country inequality increases. Therefore, median-sensitive lines are not well-suited for the evaluation of unequal growth. If the lack of robustness of the mean in random samples is judged too serious, a good compromise is to use a partial mean, such as mean income among the 99% least rich agents. For any choice of income standard, the index proposed in Section 2.4 is a strong candidate for replacing standard indices that give no priority to subsistence over social participation, such as the HC and the PGR.

2.7 Empirical illustration

In this section, I apply the new index using World Bank data. The objective is to verify that the index proposed is well-suited for evaluating unequal growth. First, using different poverty measures, I compare poverty between several low-income low-inequality countries and middle-income high-inequality countries. I show that the judgments obtained by a poverty measure based on my index are more in line with intuition than those obtained by standard measures. Second, I use the poverty measure based on my index in order to evaluate whether the economic growth taking place over the last 20 years in low- and middle-income countries was poverty reducing in spite of the increase in intra-country inequality. I discuss the variables influencing the answer.

The data is taken from PovcalNet, a website built by the World Bank that provides income and consumption data.[34] This data is gathered from more than 850 surveys of randomly sampled households in 127 low- and middle-income countries between 1981 and 2010. The frequency and precision of the surveys vary from one country to another. In some countries, the surveys focus on income, whereas in others on the value of total consumption. In order to permit cross-country comparisons, the Bank translates the survey data by making use of the Purchasing Power Parity (PPP) exchange rates for household consumption from the 2005 International Comparison Program. The national income distributions presented in PovcalNet are estimated from the survey data. More information about the data can be found in Chen and Ravallion (2013).[35]

[33]I am grateful to Karel Van den Bosch and Tim GoedemÃľ for having pointed out to me the usefulness of moving average income standards.

[34]PovcalNet: the on-line tool for poverty measurement developed by the Development Research Group of the World Bank. www.iresearch.worldbank.org/PovcalNet.

[35]PovcalNet is the database used in Chen and Ravallion (2013).

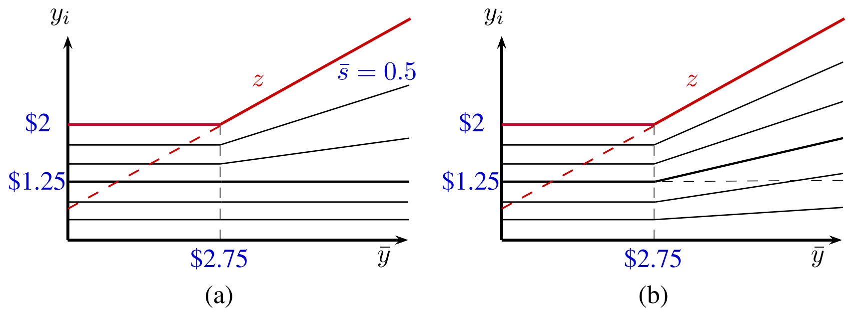

Figure 2.6: (a) Absolute-homothetic EO below the endogenous poverty line (z^a = \$1.25 a day). (b) Homothetic EO below the endogenous poverty line.

2.7.1 A poverty measure based on the new index

This section demonstrates how to apply the new index. I assume that the selected endogenous line has the following weakly relative definition, illustrated in Figure 2.6:

$$z(\overline{y}) = max(\$2, \$0.625 + 0.5\overline{y}).$$

Its income threshold equals \$2 a day in countries whose mean income is lower than \$2.75 a day. The World Bank considers that \$2 a day is the treshold for income poverty in developing countries. For mean incomes higher than \$2.75 a day, this line has a constant slope of one half. Observe that the intercept \$0.625 of this second part is positive. As a result, the line does not evolve as a constant fraction of the mean.

This line is very close to that used by Chen and Ravallion (2013). The only difference is that the income threshold for low-income countries used by these authors is \$1.25 a day, considered by the World Bank as the threshold for extreme poverty.[36] For richer countries, these authors fit their line on national thresholds. Their premise is that thresholds adopted at a country level reflect a balance made between absolute and relative aspects of income. The endogenous line selected is of course debatable but the objective pursued here is not to argue in favor of its use but rather to pick one that seems reasonable and that serves for purposes of illustration.

The new index is based on an absolute-homothetic EO below the endogenous line. The only parameter of this family of EOs is the subsistence threshold z^a. I take z^a to be the threshold for extreme poverty: \$1.25 a day. This threshold was computed as an average of income thresholds in the fifteen poorest countries of the World (Ravallion et al., 2009). Many among these countries establish their national thresholds based on the cost of a bundle of goods whose consumption guarantees to reach a minimal level of physical survival (including a minimal nutrition level). Therefore this choice seems natural for z^a. Individuals earning less than \$1.25 a day

[36] It makes little sense for my purpose to consider that agents whose income is \$1.25 a day have the same individual poverty than agents at the poverty line in richer countries since I consider \$1.25 a day to be the subsistence threshold.

are deemed absolutely poor and those earning more than \$1.25 a day but less than the endogenous threshold are deemed relatively poor. The absolute-homothetic EO defined is illustrated in Figure 2.6.a.

The poverty measure based on my index is denoted P^{EL}, where the superscript is meant to indicate that it is based on the endogenous line. Given the endogenous line selected and the choice of z^a, this poverty measure has the following mathematical expression.[37]

$$P^{EL}(y) = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{2-e^k(y_i,\overline{y})}{2}\right), \tag{2.11}$$

$$\text{where } e^k(y_i,\overline{y}) = \begin{cases} y_i & \text{if } y_i \leq 1.25, \\ 1.25 + (2-1.25)\frac{y_i - 1.25}{z(\overline{y})-1.25} & \text{otherwise.} \end{cases}$$

Judgements based on P^{EL} are compared with those obtained by four other measures. Among the four alternative measures, three are based on the Head-Count Ratio while the last is based on the Poverty Gap Ratio. The first measure, HC^{AL}, is an absolute measure corresponding to the fraction of individuals whose income is below the absolute line defined by the subsistence threshold \$1.25 a day. The second, HC^{RL}, is a relative measure corresponding to the fraction of individuals whose income is below the relative line whose threshold is half the mean income. This measure provides some information about the inequality in the distribution. The third measure, HC^{EL}, is an endogenous measure corresponding to the fraction of individuals whose income is below the endogenous line defined above. The last measure, PGR^{EL}, is the Poverty Gap Ratio below the endogenous line, defined by:

$$PGR^{EL}(y) = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z(\overline{y})-y_i}{z(\overline{y})}\right).$$

This last measure satisfies both *Monotonicity in Income* and *Transfer among Poor* but gives no priority to subsistence over social participation.

I now consider the relations existing between P^{EL} and PGR^{EL}. For mean incomes below \$2.75 a day, the endogenous line is flat. The respective EOs of P^{EL} and PGR^{EL}, illustrated respectively in Figure 2.6.a and 2.6.b, are hence equivalent for these low values of mean incomes. As a result, P^{EL} and PGR^{EL} return equal values for very poor countries. Above \$2.75 a day, P^{EL} systematically returns lower figures than PGR^{EL} because the absolute-homothetic EO of P^{EL} associates to any bundle an equivalent income at \$2.75 a day larger than the one associated to the same bundle by the homothetic EO of PGR^{EL}. Therefore, if distribution A has a larger mean income than distribution B with $\overline{y}^B = 2.75$ and P^{EL} concludes that there is more poverty in A than in B, then PGR^{EL} draws the same conclusion. Index PGR^{EL} places more emphasis on poverty in richer countries as its homothetic

[37] Given that the endogenous line is flat for mean incomes below $y^k = 2.75$, the PGR at the origin is equivalent to the PGR at y^k.

EO weighs more the relative aspect of individual poverty.

2.7.2 Empirical results

The data extracted from PovcalNet is used for computing the five poverty measures. I first show that, when dealing with unequal growth, P^{EL} makes poverty judgments that are more in line with intuition than those of the other four measures.

Table 2.4 provides figures for six countries in 2010. The countries are sorted in increasing order of mean income. Three low-income low-inequality countries are considered, namely Ethiopia, Nepal and Ivory Coast. Their mean incomes amount to \$2, \$2.2 and \$3 a day respectively and their Gini coefficients in 2010 amount to 34%, 33% and 43%.[38] Three middle-income high-inequality countries are considered, namely Bolivia, South Africa and Brazil. Their mean incomes amount to \$8.3, \$8.4 and \$13.8 a day respectively and their Gini coefficients in 2010 amount to 50%, 63% and 54%. Remember that for my purpose, the distributions of two countries can equally be interpreted as two distributions corresponding to the same country but at different points in time.

Table 2.4: Cross-country comparison of poverty figures in 2010.

Countries	Mean	Gini	HC^{AL}	HC^{RL}	HC^{EL}	PGR^{EL}	P^{EL}
Ethiopia	2.0	34	30.6	17.7	65.0	23.1	23.1
Nepal	2.2	33	24.8	18.5	56.3	18.7	18.7
Ivory Coast	3.0	43	22.7	30.0	47.6	18.3	17.4
Bolivia	8.3	50	13.4	43.3	48.3	25.3	16.5
South Africa	8.4	63	13.8	57.1	61.3	32.8	17.6
Brazil	13.8	54	5.4	43.1	46.5	22.1	11.7

All poverty measures and the Gini coefficients are expressed in %. Mean incomes are expressed in \$ a day (2005 PPP). Source: PovcalNet.

HC^{AL} is strongly negatively correlated with mean income and HC^{RL} is strongly positively correlated with inequality, as measured by the Gini coefficient.[39] HC^{RL} concludes that middle-income countries, having a larger income inequality, have by far the largest poverty. HC^{AL} reaches opposite conclusion. On the sole basis of these two measures, it is hence difficult to balance the absolute and relative aspects of growth. The three measures based on the endogenous line are more nuanced. PGR^{EL} places more emphasis on poverty in richer countries and concludes that the two poorest countries are Bolivia and South Africa. In contrast, the two poorest countries according to P^{EL} are low-income countries, namely Ethiopia and Nepal.

[38] The Gini coefficient is a popular measure of inequality. The larger the Gini coefficient, the larger is inequality. The figures were obtained online from the World Bank Poverty and Equity Database on the 24^{th} of August 2015, www.povertydata.worldbank.org. The Gini coefficient is measured in 2010 for Ethiopia and Nepal; in 2009 for Bolivia, South Africa and Brazil and in 2008 for Ivory Coast.

[39] In the sample, the coefficients of correlations are -0.97 and 0.99 respectively.

Pairwise comparisons of countries having different mean incomes illustrate the different judgments made by P^{EL}, PGR^{EL} and HC^{EL}. PGR^{EL} and HC^{EL} conclude that there is less – or approximately equal – poverty in Ivory Coast than in Brazil, even if the fraction of absolutely poor individuals is much higher in the former (22.7 %) than in the latter (5.4%). In contrast, P^{EL} places more emphasis on the absolute aspects of income poverty and concludes that there is more poverty in Ivory Coast than in Brazil. Oppositions of the same type can be found when comparing South Africa with Nepal or Ivory Coast, or when comparing Brazil with Bolivia. Observe that P^{EL} does not always follow the judgments of HC^{AL}. Unlike P^{EL}, HC^{AL} concludes that there is much more poverty in Nepal and Ivory Coast than in South Africa. If Nepal and Ivory Coast underwent a very unequal growth transforming their distributions into that of South Africa, whose distribution is very polarized, P^{EL} would not lead to conclusions as enthusiastic as those obtained from HC^{AL}. Observe that the difference in judgments described above are based on large differences in the respective figures.

Table 2.4 demonstrates that the poverty judgments drawn from P^{EL} can be radically different from those obtained with the other four measures. Moreover, the judgments drawn from P^{EL} seem to be in line with basic intuitions. Next, P^{EL} is used in order to evaluate the impact of the unequal growth taking place over the period 1990-2010 in different geographic entities. The decomposability of P^{EL} allows us to analyze the variables influencing the poverty judgments.

Table 2.5: Evaluation of several unequal growths.

Geo Entity	Year	Mean	HC^{RL}	HC^{AL}	HC^{EL}	P^{EL}	P^a/P^{EL}
World	1990	3.0	21.2	43.0	70.7	30.7	0.82
	2010	4.9	26.4	20.8	52.7	17.7	0.66
Urban China	1990	1.9	9.1	23.4	61.2	18.9	0.62
	2010	7.1	21.7	0.6	30.6	4.7	0.08
Costa Rica	1990	7.0	31.5	8.4	40.0	11.4	0.53
	2010	15.3	40.3	2.6	43.7	8.9	0.22
Mexico	1990	7.8	24.1	4.5	29.2	7.4	0.39
	2010	10.6	35.8	0.7	41.2	7.5	0.05
Hungary	1996	8.8	9.8	0.2	16.0	1.7	0.07
	2010	12.5	15.2	0.2	20.1	2.2	0.06

All poverty measures are expressed in %. Mean income is expressed in \$ a day (2005 PPP). P^a corresponds to the contribution to P^{EL} of absolutely poor agents, defined by (2.7). Source: PovcalNet.

Table 2.5 provides the before- and after-growth figures for five geographic entities.[40] All five geographic entities experienced an increase in mean income together with an increase in inequality, as indicated by HC^{RL}. P^{EL} allows us to decompose

[40]The figures for the World are an aggregate of the figures for the low- and middle-income countries, weighted by their population. The figures for urban China are obtained by computing the endogenous threshold for the mean income in urban China.

the fraction of poor individuals (HC^{EL}) between those that are absolutely poor (HC^{AL}) and those that are "only" relatively poor. Furthermore, the figure for P^{EL} can be decomposed between the contribution of absolutely poor agents (P^a) and that of relatively poor agents (P^r). These decompositions are illustrated in Figure 2.7 for the World, urban China and Mexico. In Figure 2.7, the contribution of absolutely poor agents (P^a) is further decomposed between its absolute (P^{aa}) and relative components (P^{ar}).[41]

The World and urban China experienced a large decline in income poverty over the period: P^{EL} dropped by 42% and 75%, respectively. In other words, in spite of the increase in income inequality, particularly important in urban China as indicated by HC^{RL}, P^{EL} concludes unambiguously that growth has been poverty reducing.[42] These reductions reflect primarily the changes in absolute poverty. Absolute poverty was a main concern in both entities in 1990. In the World for example, 43% of individuals were absolutely poor in 1990 and these individuals contributed to 82% of P^{EL}. In 2010, only 20.8% of individuals remainded absolutely poor in the World, contributing then to 66% of P^{EL}. For urban China, absolute poverty has been almost eradicated over the period. These evolutions and trade-offs appear clearly when studying the graphs decomposing P^{EL} in Figure 2.7. The decrease in P^{EL} in both entities is clearly driven by changes in P^a whereas at the same time P^r does not change much.

Costa Rica and Mexico experienced a lower reduction in poverty than the World and urban China over this period. P^{EL} dropped by 22% in Costa Rica whereas it returned to its initial value in Mexico. The increase in relative poverty mitigated the significant reduction in absolute poverty achieved by the two countries. Absolute poverty was an important concern in 1990 – 53% of P^{EL} for Costa Rica and 39% of P^{EL} for Mexico – although not as dominant as for the World and urban China. The fraction of absolutely poor individuals fell from 8.4% to 2.6% in Costa Rica and from 4.5% to 0.7% in Mexico. At the same time however, the large increase in inequality in these two countries implied that more individuals were poor in 2010 than in 1990, as shown by HC^{EL}. Again, the trade-offs for Mexico appear clearly when studying the graphs decomposing P^{EL} in Figure 2.7. In Mexico, the large increase in inequality taking place between 1990 and 1994 increased significantly P^r. The later reduction in P^a only compensated for the increase in P^r. Appendix 2.9.10 contains a further analysis of the Mexican case based on cumulative distributions of income and individual poverty.

Hungary experienced an increase in poverty over the period 1996 – 2010, in spite of an increase of 43% of its mean income. P^{EL} increased by 30% in Hungary over the period. Absolute poverty was not an important concern in 1996 – P^a was less than 10 % of P^{EL} in 1996 – and did not change significantly over the period. On the contrary, income inequality increased and 20% of individuals were poor in 2010 whereas only 16 % of individuals were poor in 1996. The increase in P^{EL} is directly

[41] P^{aa} and P^{ar} correspond respectively to term 1 and term 2 in (2.7).

[42] It is the *intra-country* inequality that is accounted for when discussing the evolution of inequality in the World. Intra-country inequality influences poverty in the World via its impact on the country-specific endogenous thresholds.

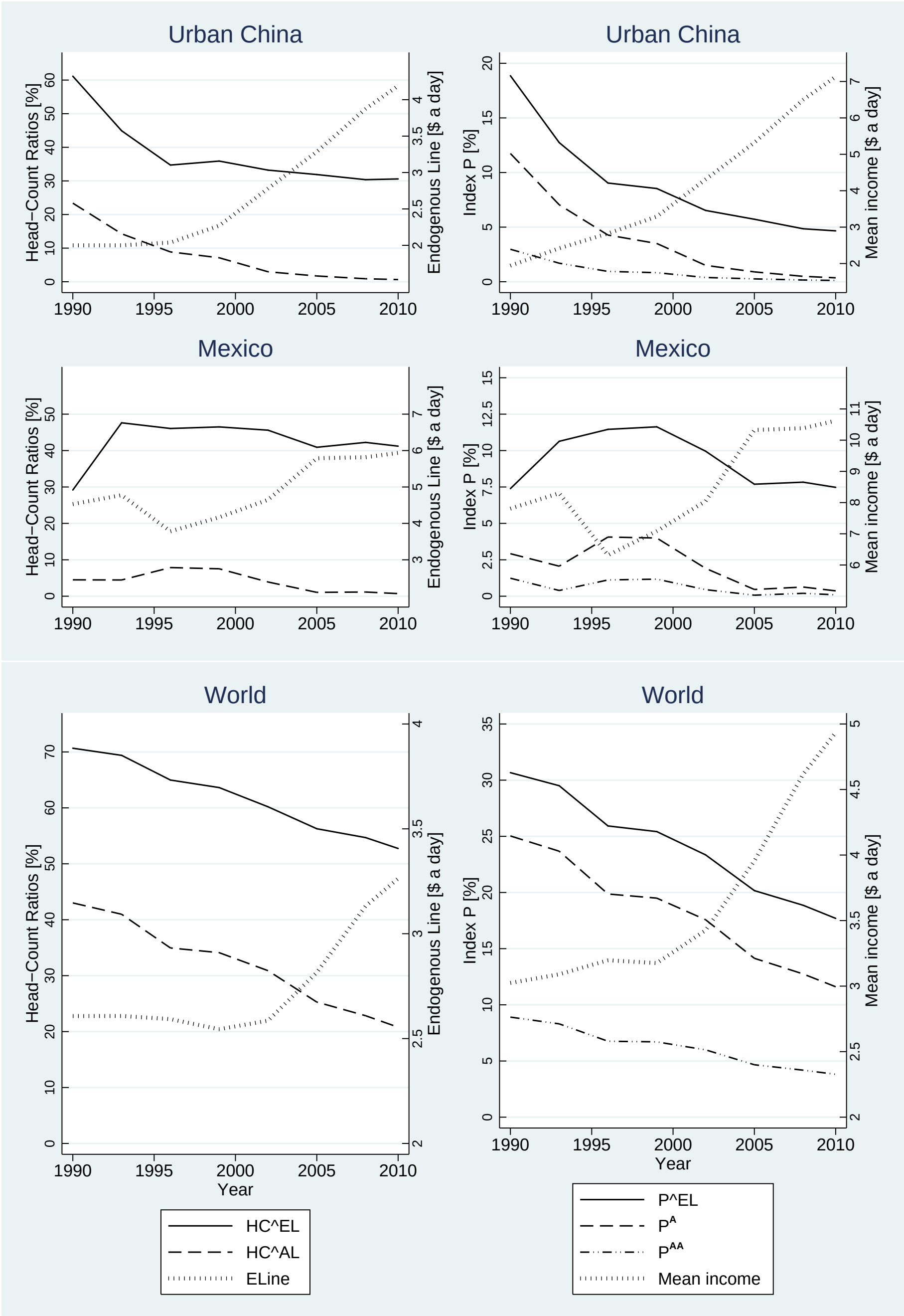

Figure 2.7: Evolution of income poverty between 1990 and 2010 in urban China, Mexico and the World as measured by P^{EL}. The left graphs show the decomposition of poor agents (HC^{EL}) between absolutely poor (HC^{AL}) and relatively poor, together with the endogenous threshold. The right graphs show the decomposition of P^{EL} between the contribution of absolutely poor agents (P^a) and that of relatively poor agents (P^r), together with mean income. P^a is further decomposed between its absolute (P^{aa}) and relative contributions (P^{ar}). Source: PovcalNet.

driven by the increase in P^r.

The distinction between absolutely and relatively poor agents and the decomposability of the index make it possible to separately track these two forms of poverty and aggregate them in a coherent way. I illustrate this possibility for the case of urban China, shown in Figure 2.7. In urban China in 1990, 23.4% of individuals were absolutely poor and 37.8% were relatively poor, adding up to 61.2% of poor individuals. Overall, the poverty index for the income distribution in 1990 takes a value of 18.9%. This value of income poverty can be decomposed into the contribution of absolutely poor agents (11.8%) and that of relatively poor agens (7.1%). Hence, absolutely poor agents contributed to 62% of income poverty, which shows that absolute poverty was the main issue in urban China in 1990. In 2010, 0.6% of individuals were absolutely poor and 30% were relatively poor, adding up to 30.6% of poor individuals. Overall, the poverty index for the income distribution in 2010 takes a value of 4.7%, a figure 75% lower than that of 1990. This lower value of income poverty can be decomposed into the contribution of absolutely poor agents (0.4%) and that of relatively poor agens (4.3%). Hence, absolutely poor agents contributed to 8% of income poverty. This demonstrates that the reduction in absolute poverty is responsible for most of this three-quarters reduction in income poverty. Moreover, it shows that relative poverty became the main issue in urban China in 2010.

Analyzing with P^{EL} several unequal growths has shown that very different conclusions can be drawned by this measure. Different factors influence the conclusions of P^{EL}, such as the extent of growth or the extent of the increase in inequality. A key factor is the importance for P^{EL} of absolute poverty at the beginning of the period. If absolute poverty is not the main concern, like in Hungary, the increase in inequality entails an increase in P^{EL}.

Altogether, P^{EL} confirms that poverty reduction has been impressive over the last decades in low- and middle-income countries ("the World" in Table 2.5). In fact, poverty decreased even more than Head-Count based measures suggest. Over the period 1990-2010, even if the fraction of poor individuals decreased only by 25% , P^{EL} concludes that income poverty was reduced by 42%.

2.8 Concluding remarks

Comparing income poverty between societies with different standards of living has always been done with extreme caution. This caution follows in part from the inability of standard poverty measures to consider simultaneously the absolute and relative aspects of income poverty. Bringing together the concepts of endogenous lines (Foster, 1998; Ravallion and Chen, 2011) and other-regarding preferences, I show how these aspects can be combined by endogenous poverty measures based on a new index; therefore providing a firmer foundation for these comparisons.

The distrust of standard poverty measures has complicated the evaluation of unequal growth. A literature proposing several definitions for *pro-poor growth* emerged in order to fill the gap.[43] Araar and Duclos (2009) classified the different proposals in two categories, the absolute and the relative pro-poorness measures. The existence of these two categories shows that the pro-poor growth literature is confronted to the difficulty of considering simultaneously the absolute and relative aspects of income. My index constitutes a possible answer to this difficulty. In the spirit of Ravallion and Chen (2003), growth could be deemed pro-poor if it leads to a decrease in an endogenous poverty measure based on my index. The endogenous line and the subsistence threshold become then the key parameters for the evaluation of the pro-poorness of growth.

There are several direct applications for this research. A proeminent example is the measurement of income poverty by the World Bank. This institution recently established a commission aimed at advicing it on the best way to monitor the realization of its twin goals.[44] The decomposition of the new index between absolute and relative poverty should simplify the analysis and the communication on the progress achieved towards its twin goals. In the same vein, the EU Commission could replace the AROPE measure by a measure based on the new index. Countries whose official income poverty definition is judged non-satisfactory could also use the new index, especially if they experience unequal growth. The United States constitute a prominent example as several observers like Ruggles (1990) and Citro and Michael (1995) questioned its absolute line. See Blank (2008) for a review of the political initiatives that have attempted to modify it.

Switching the poverty measure changes the evaluation of policies aimed at reducing poverty. Up to now, policy makers used absolute measures for policy evaluation in low- and middle-income countries and relative measures in high-income countries. This practice ensures that the most relevant aspect of income poverty is captured in each case, at the cost of ignoring the other aspect. The limitation of this practice is that it yields extreme judgments on growth. On the one hand, absolute measures evaluate policies creating economic growth positively, regardless of their distributional aspects. On the other hand, relative measures judge redistributive policies positively, regardless of their impact on growth, as long as the inequality experienced by the poor decreases. The evaluation of policies with a measure based on the new index solves these limitations. This index combines both aspects and emphasizes more the aspect that is dominant in the distribution considered. Indeed, its judgments depend on the importance of absolute poverty in the initial distribution. As a consequence, the policies recommended by this index should be in line with what

[43] A basic definition of pro-poorness is to require that growth reduces a poverty measure based on the Watts index (Ravallion and Chen, 2003). This definition has been called "weak" as it does not specify a minimal extent of poverty reduction for a given growth in mean income (Kakwani, 2008). Alternatively, growth can be deemed pro-poor if the average growth among the poor is higher than the growth in mean income (Duclos, 2003). Another contribution from Foster and Szekely (2008) aims at getting around the arbitrariness inherent in a poverty line. These authors suggest comparing the growth rate in mean income with that of different generalized means. The lower the parameter β defining a particular generalized mean, the more emphasize is put on incomes at the bottom of the distribution.

[44] The Commission on Global Poverty was established in 2015. http://www.worldbank.org/en/programs/commission-on-global-poverty.

the specific situation requires.

The index proposed has applications outside income poverty measurement. If the emphasis has been put on income, the index can measure the poverty in any other resource for which both the absolute and relative aspects matter, like education or health.

More generally, this research contributes to attempts at introducing relative considerations into the normative evaluation of economic outcomes. Adapting utilitarian indicators to other-regarding preferences is of course straightforward. Nevertheless, utilitarian indicators often provide judgments that are at odds with equality of opportunity principles. Social ordering functions, i.e. indicators of well-being derived from efficiency and fairness principles, offer in that respect a good alternative to utilitarian indicators (Fleurbaey and Schokkaert, 2011). Recently, a nascent literature has started investigating how to derive social ordering functions for economies populated with other-regarding agents. See Treibich (2014) for the single-good case and Decerf and Van der Linden (2014) for the multi-good case.

2.9 Appendix

2.9.1 Proof of Theorem 1

I show that statement 2 implies statement 1. Take any endogenous line z and any poverty index P satisfying the five axioms.

STEP 1: From a poverty ordering on income distributions to a poverty ordering on distributions of individual poverty.

I define a continuous mapping $m : Y \to \mathbb{R}^{N'}$, where $N' := \{n \in \mathbb{N} | n \geq 2\}$. Let $\succeq$ be a EO in $\mathcal{R}$ whose unanymous judgments among the poor are respected by P. By *Domination among Poor*, such $\succeq$ exists. Consider any numerical representation d of $\succeq$. For each $(y_i, \overline{y}) \in X$, let $\nu_i := d(y_i, \overline{y})$. Mapping m is defined for all $y \in Y$ such that

$$m(y) = (\nu_1, \cdots, \nu_{n-1}) := \nu.$$

Observe that if distribution y has n components, then $m(y)$ has $n-1$ components. The size of distribution ν is taken to be $n-1$ as for all $y \in Y$ we have $d(y_n, \overline{y}) = 0$ since $y_n \geq z(\overline{y})$ and is hence omitted. Mapping m is continuous since d is continuous in both its arguments and the mean is a continuous function of its arguments. Given the numerical representation d, mapping m returns the distribution of individual poverties corresponding to any income distribution.

I show for the mapping defined that $m(Y) = V_d := [0, 1]^{N'}$. The domain of images of Y through mapping m is hence a product space: $V_d = \times_{i=1}^{N'}[0, 1]_i$. This means that (i) $m(Y) \subseteq V_d$ and (ii) $V_d \subseteq m(Y)$, that is for all $\nu \in V_d$ there exists $y \in Y$ such that $m(y) = \nu$. If (i) follows directly from the definition of mapping m, (ii) remains to be proven. Lemma 5 proves that $V_d \subseteq m(Y)$.

Lemma 5. *For all endogenous line z, $\succeq \in \mathcal{R}$ and $\nu \in V_d$, there exists $y \in Y$ such that $\nu = m(y)$.*

Proof. Take any endogenous line z, $\succeq \in \mathcal{R}$ and $\nu \in V_d$. Let $g > 0$ be such that $g \geq z(g)$. Such g always exists by restriction Possibility of Poverty Eradication. We construct y such that $\overline{y} = g$ and $m(y) = \nu$. For all $i \leq q$, y is such that $y_i := a_i$ defined implicitly by $\nu_i = d(a_i, g)$. By restriction Minimal Absolute Concern and the continuity of d, we have that $a_i \in [0, z(g))$ for all $i \leq q$. Let y' be such that $y'_i := y_i$ for all $i \leq q$ and $y'_j := g$ for all j with $q+1 \leq j \leq n$. We have $\overline{y}' \leq g$ as $z(g) \leq g$. There exists hence $\ell \geq g$ such that, if $y_j := \ell$ for all j with $q+1 \leq j \leq n$, then we have $\overline{y} = g$. As $\ell \geq g \geq z(g)$, all agents j with $q+1 \leq j \leq n$ are non-poor. By construction we have $m(y) = \nu$. ∎

P is by definition the representation of a complete poverty ordering $\succeq_Y$ on Y. By *Domination among Poor*, for any two $y, y' \in Y$ such that $m(y) = m(y')$ we have $P(y) = P(y')$. Therefore, the complete ordering $\succeq_Y$ implies a complete ordering $\succeq_{V_d}$ on V_d since $V_d = m(Y)$. Ordering $\succeq_{V_d}$ is defined such that for all $y, y' \in Y$ we have $y \succeq_Y y' \Leftrightarrow m(y) \succeq_{V_d} m(y')$. Ordering $\succeq_{V_d}$ is continuous since the ordering on Y is continuous by *Continuity* and mapping m is continuous. Being continuous, ordering $\succeq_{V_d}$ can be represented by a continuous index $P^\nu : V_d \to \mathbb{R}$. In particular, ordering $\succeq_{V_d}$ is represented by P^ν defined such that for all $\nu \in V_d$ and $y \in Y$ with $m(y) = \nu$, we have $P^\nu(\nu) = P^\nu\big(m(y)\big) = P(y)$.

STEP 2: Index P^ν representing ordering $\succeq_{V_d}$ on distributions of individual poverty is additively separable.

If the assumptions of Theorem 1 in Gorman (1968) are all met, then for any $n \in N$ and any ν of size $n-1$, index P^ν has the following functional form:

$$P^\nu(\nu) = \tilde{F}\left(\sum_{i=1}^{n-1} \tilde{\varphi}(\nu_i)\right) \tag{2.12}$$

where $\tilde{F}$ and $\tilde{\varphi}$ are strictly increasing functions.

Take any $n \in N$. For the remaining part of Step 2, I abuse slightly notation by denoting V_d the subset of V_d containing elements of size $n-1$. The three assumptions required for this theorem are the following:

Assumption 1: There exists a *complete* and *continuous* ordering on a *product space.*

I proved in Step 1 that the ordering $\succeq_{V_d}$ is complete and continuous on V_d, which is a product space $V_d = \times_{i=1}^{n-1}[0,1]_i$.

Assumption 2: Each sector $[0,1]_i$ of V_d has a *countably dense subset*, is *arc-connected* and is *strictly essential.* *Strict essentiality* means that given any sub-distribution $(\nu_1, \cdots, \nu_{i-1}, \nu_{i+1}, \cdots, \nu_{n-1}) \in \times_{j=1}^{n-2}[0,1]_j$, not all elements of $[0,1]_i$ are indifferent for the ordering $\succeq_{V_d}$.

As all sectors are real intervals. Any sector therefore has a *countably dense subset* and is *arc-connected*. Strict essentiality follows directly from *Domination among Poor* together with the fact that for any $i \leq n-1$ and any subdistribution $(\nu_1, \cdots, \nu_{i-1}, \nu_{i+1}, \cdots, \nu_{n-1})$, the individual poverty ν_i is not constrained as the individual poverty of the agent n with highest income is discarded.[45]

Assumption 3: Let $S := \{[0,1]_1, \cdots, [0,1]_{n-1}\}$ be the set of sectors in V_d and $A \subseteq S$ be any subset of sectors, we have that each A is *separable*. *Separability* means that for all $(u,w),(v,w),(u,t),(v,t) \in V_d$, we have $P^\nu(u,w) \geq P^\nu(v,w) \Leftrightarrow P^\nu(u,t) \geq P^\nu(v,t)$. Separability is proven in two substeps.

Substep 1: Construct for each of the four distributions of individual poverty (u,w), (v,w), (u,t) and (v,t) a particular income distribution associated to it.

Construct $y^1, y^2, y^3, y^4 \in Y$ such that $m(y^1) = (u,w)$, $m(y^2) = (v,w)$, $m(y^3) = (u,t)$, $m(y^4) = (v,t)$ and $\overline{y}^1 = \overline{y}^2 = \overline{y}^3 = \overline{y}^4 = g$ with $g \geq z(g)$. Such distributions exist and are constructed following the procedure given in Lemma 5.

Decompose in subgroups $y^1 = (y^1_A, y^1_B, y^1_n)$, such that subdistributions y^1_A and y^1_B are associated – via the numerical representation d – to the subdistributions u and w respectively.[46] Typically, $\overline{y}^1_A \neq \overline{y}^1_B \neq g$ but the next operations aims at obtaining such equality.

Triplicate y^1 and re-organize the subgroups to obtain at least one non-poor agent per subgroup. Let $y^{1'} := (y^1, y^1, y^1) = (y^1_A, y^1_A, y^1_A, y^1_B, y^1_B, y^1_B, y^1_n, y^1_n, y^1_n)$. This triplication does not affect the mean: $\overline{y}^{1'} = \overline{y}^1$. Reorganize subgroups: $y^{1'} = (y^1_{A'}, y^1_{B'}, y^1_n)$ with $y^1_{A'} := (y^1_A, y^1_A, y^1_A, y^1_n)$ and $y^1_{B'} := (y^1_B, y^1_B, y^1_B, y^1_n)$. Letting $u' := (u,u,u)$ and $w' := (w,w,w)$, we have that

$$m(y^{1'}) = (u,u,u,0,w,w,w,0) = (u',0,w',0),$$

as $d(y_i, g) = 0$ for any $y_i \geq z(g)$.

Construct $y^{1*}_{A'}$ such that $m(y^{1*}_{A'}) = u'$ with $\overline{y}^{1*}_{A'} = g$ and $y^{1*}_{B'}$ such that $m(y^{1*}_{B'}) = w'$ with $\overline{y}^{1*}_{B'} = g$. Those income distributions exist as proven in Lemma 5, as both subgroups A' and B' contain at least one non-poor agent. The income distribution $y^{1*} := (y^{1*}_{A'}, y^{1*}_{B'}, g)$ is such that $m(y^{1*}) = (u',0,w',0)$. This distribution is such that $\overline{y}^{1*} = g$ as its three subgroups have mean g.

Using the same procedure (decomposition, triplication, reorganization), construct successively $y^{2'}, y^{3'}, y^{4'}$ and y^{2*}, y^{3*}, y^{4*} such that:

$$\begin{aligned}
y^{1*} &= (y^{1*}_{A'}, y^{1*}_{B'}, g) && \text{with } m(y^{1*}) = (u',0,w',0) = (u,u,u,0,w,w,w,0),\\
y^{2*} &= (y^{2*}_{A'}, y^{2*}_{B'}, g) && \text{with } m(y^{2*}) = (v',0,w',0) = (v,v,v,0,w,w,w,0),\\
y^{3*} &= (y^{3*}_{A'}, y^{3*}_{B'}, g) && \text{with } m(y^{3*}) = (u',0,t',0) = (u,u,u,0,t,t,t,0),\\
y^{4*} &= (y^{4*}_{A'}, y^{4*}_{B'}, g) && \text{with } m(y^{4*}) = (v',0,t',0) = (v,v,v,0,t,t,t,0).
\end{aligned}$$

For all $m \in \{1,2,3,4\}$, we have $P(y^{m'}) = P(y^m)$ by *Replication Invariance*. As

[45] In the definition and the proof of strict essentiality, the indices are not sorted by income level but refer to the identities.

[46] For each element $u_i \in u$ there exists $y^1_i \in y^1_A$ such that $u_i = d(y^1_i, \overline{y}^1)$. The same holds for w and y^1_B.

$(y_i^{m'}, g) \sim (y_i^{m*}, g)$ for all $i \leq q(y^{m*})$, we have $P(y^{m*}) = P(y^{m'})$ by *Domination among Poor*. Therefore, proving $P(y^{1*}) \geq P(y^{2*}) \Leftrightarrow P(y^{3*}) \geq P(y^{4*})$ is equivalent to proving $P^{\nu}(u, w) \geq P^{\nu}(v, w) \ \Leftrightarrow \ P^{\nu}(u, t) \geq P^{\nu}(v, t)$. For notational simplicity, drop the symbols * and $'$ to name the new distributions and subgroups as the old ones.

Substep 2: Prove separability from judgments on the associated income distributions: $P(y_A^1, y_B^1, g) \geq P(y_A^2, y_B^2, g) \Leftrightarrow P(y_A^3, y_B^3, g) \geq P(y_A^4, y_B^4, g)$.

These income distributions are constructed such that $P(y_A^1) = P(y_A^3)$, $P(y_A^2) = P(y_A^4)$, $P(y_B^1) = P(y_B^2)$ and $P(y_B^3) = P(y_B^4)$ by *Domination among Poor*. By assumption, we have $P(y^1) \geq P(y^2)$. As $P(y_B^1) = P(y_B^2)$, we have that $P(y_A^1, g) \geq P(y_A^2, g)$ by *Weak Subgroup Consistency* (remember all our subgroups have their mean equal to g). By *Weak Subgroup Consistency* again, this implies $P(y_A^1) \geq P(y_A^2)$.[47]

Then, $P(y_A^1) \geq P(y_A^2)$ together with $P(y_A^1) = P(y_A^3)$ and $P(y_A^2) = P(y_A^4)$ imply $P(y_A^3) \geq P(y_A^4)$. Two cases can arise.

- Case 1: $P(y_A^3) > P(y_A^4)$.
  As $P(y_B^3) = P(y_B^4)$, we have $P(y_B^3, g) = P(y_B^4, g)$ by *Domination among Poor*. Together we obtain $P(y_A^3, y_B^3, g) > P(y_A^4, y_B^4, g)$ by *Weak Subgroup Consistency*. This case is hence such that $P(y^3) \geq P(y^4)$, as desired.

- Case 2: $P(y_A^3) = P(y_A^4)$.
  I show by contradiction this case is such that $P(y^3) \geq P(y^4)$. Assume we have $P(y_A^3, y_B^3, g) < P(y_A^4, y_B^4, g)$. As $P(y_A^3) = P(y_A^4)$, *Weak Subgroup Consistency* implies that $P(y_A^3, y_B^3, y_A^4, g) < P(y_A^4, y_B^4, y_A^3, g)$. Again, as $P(y_B^3) = P(y_B^4)$, we obtain $P(y_A^3, y_B^3, y_A^4, y_B^4, g) < P(y_A^4, y_B^4, y_A^3, y_B^3, g)$. This is a contradiction as the two distributions have equal poverty by *Symmetry*.

The two cases lead to $P(y^3) \geq P(y^4)$, which proves separability.

As all three assumptions hold, we can use Theorem 1 in Gorman (1968) and obtain, for all $\nu \in V_d$:

$$P^{\nu}(\nu) = \tilde{F}'\left(\sum_{i=1}^{n-1} \tilde{\varphi}_i(\nu_i)\right)$$

where $\tilde{F}'$ and $\tilde{\varphi}_i$ are strictly increasing functions. Functions $\tilde{\varphi}_i$ might still depend on the rank i of the considered agent. Nevertheless, since $\succeq_{V_d}$ is separable, we must have $\tilde{\varphi}_i = \tilde{\varphi} + f(i)$. Defining $\tilde{F}(x) := \tilde{F}'(x + \sum f(i))$, a translation of $\tilde{F}'$, we can use (2.33) with function $\tilde{\varphi}$ independent of rank i.

STEP 3: Show functions $\tilde{F}$ and $\tilde{\varphi}$ do not depend on the number n of agents.

[47] Strictly speaking *Weak Subgroup Consistency* cannot be applied again as subgroup g contains a unique agent and hence does not belong to Y. Nevertheless, further replications of the income distributions solve the issue.

Theorem 1 in Gorman (1968) is valid for a fixed number n of agents. Therefore, when n is allowed to vary, equation (2.33) must be written:

$$P^{\nu}(\nu) = \tilde{F}_n\left(\sum_{i=1}^{n-1} \tilde{\varphi}_n(\nu_i)\right).$$

I modify the proof of Foster and Shorrocks (1991) in order to show that these functions are independent of n.

Step 3.1: Define transformations of $\tilde{F}_n$ and $\tilde{\varphi}_n$ for normalization purposes.

Let F_n and φ_n be the following transformations of $\tilde{F}_n$ and $\tilde{\varphi}_n$:

$$\varphi_n(\nu_i) = n\left[\tilde{\varphi}_n(\nu_i) - \tilde{\varphi}_n(0)\right],$$
$$F_n(x) = \tilde{F}_n\left[x + (n-1)\tilde{\varphi}_n(0)\right].$$

These transformations allows rewritting last equation in the following way

$$P^{\nu}(\nu) = F_n\left(\frac{1}{n}\sum_{i=1}^{n-1} \varphi_n(\nu_i)\right),$$

where $\varphi_n(0) = 0$.

Since agent n is non-poor by definition, we have $d(y_n, \overline{y}) = 0$. Therefore, we obtain – slightly abusing notation (by introducing agent n's zero individual poverty at the end of distribution ν) – that for all $n \geq 3$:

$$P^{\nu}(\nu) = F_n\left(\frac{1}{n}\left(\varphi_n(0) + \sum_{i=1}^{n-1} \varphi_n(\nu_i)\right)\right) = F_n\left(\frac{1}{n}\sum_{i=1}^{n} \varphi_n(\nu_i)\right), \tag{2.13}$$

where F_n and φ_n are continuous, strictly increasing and $\varphi_n(0) = 0$.

Step 3.2: Use *Replication Invariance* to prove functions F_n and φ_n do not depend on n.

From the previous step, we have $\varphi_n : [0,1] \to [0, b_n]$ with $\varphi_n(0) = 0$ for all $n \in \mathbb{N}_{++}$. Take any $y \in Y$ with dimension $n = 5$ such that a single agent is poor in y. Consider $x := (y, \ldots, y)$ a k-replication of y. Let $\nu := m(y) = (t, 0, 0, 0)$ be the individual poverty distribution associated to y where t can be any element in $[0, 1]$. Let $\nu' := m(x) = (t, \ldots, t, 0, \ldots, 0)$ be the individual poverty distribution associated to x which contains $4k - 1$ zeros and k t's. The dimension of ν is $r = 4$ and the dimension of ν' is $s = 5k - 1$. Therefore we have $s = k(r + 1) - 1 = kr + k - 1$.

Denoting $F := F_4$ and $\varphi := \varphi_4$, the relationship between φ, φ_s, F and F_s for all $t \in [0, 1]$ is computed using (2.13) and *Replication Invariance*:

$$P^{\nu}(\nu) = F\left[\frac{1}{5}\varphi(t)\right] = F_s\left[\frac{k}{5k}\varphi_s(t)\right] = P^{\nu}(\nu'),$$

$$\varphi_s(t) = 5F_s^{-1}\left[F(\frac{1}{5}\varphi(t))\right]. \tag{2.14}$$

Replacing $\varphi_s(t)$ in (2.13) by its value obtained in (2.14), we get:

$$F^{-1}[P^\nu(\nu')] = F^{-1}\left[F_s\left(\frac{1}{5k}\sum_{i=1}^{5k} 5F_s^{-1}\left[F\left(\frac{1}{5}\varphi(\nu_i')\right)\right]\right)\right] \tag{2.15}$$

$$= G_s^{-1}\left(\frac{1}{5k}\sum_{i=1}^{5k} 5G_s\left(\frac{1}{5}\varphi(\nu_i')\right)\right), \tag{2.16}$$

where $G_s(w) := F_s^{-1}(F(w))$ and $G_4(w) = F^{-1}(F(w)) = w$.

By *Replication Invariance*, we have that $F^{-1}[P^\nu(\nu)] = F^{-1}[P^\nu(\nu')]$, which by (2.16) yields:

$$G_s\left(\frac{1}{5}\varphi(t)\right) = \left(\frac{1}{5k}\sum_{i=1}^{5k} 5G_s\left(\frac{1}{5}\varphi(\nu_i')\right)\right)$$

$$= G_s\left(\frac{1}{5}\varphi(t)\right) + \frac{4k-1}{k}G_s(0),$$

which shows that $G_s(0) = 0$.

Consider now any $y' \in Y$ with dimension $n = 5$ such that two agents are poor in y'. Consider $x' := (y', \dots, y')$ a k-replication of y'. Let $\nu := m(y') = (t, u, 0, 0)$ be the individual poverty distribution associated to y' where t and u can be any element in $[0, 1]$. Let $\nu' := m(x') = (t, \dots, t, u, \dots, u, 0, \dots, 0)$ be the individual poverty distribution associated to x which contains $3k - 1$ zeros, k t's and k u's.

By *Replication Invariance*, we have that $F^{-1}[P^\nu(\nu)] = F^{-1}[P^\nu(\nu')]$, which by (2.16) yields:

$$\frac{1}{5}\varphi(t) + \frac{1}{5}\varphi(u) = G_s^{-1}\left(G_s\left(\frac{1}{5}\varphi(t)\right) + G_s\left(\frac{1}{5}\varphi(u)\right)\right),$$

which can be rewritten as the Jensen equation:

$$G_s(x + x') = G_s(x) + G_s(x'),$$

that admits as general solution $G_s(x) = r_s x + q_s$. As $G_s(0) = 0$ we have $q_s = 0$.

Replacing G_s by its expression in (2.16), we obtain

$$F^{-1}[P^\nu(\nu')] = \frac{1}{5k}\sum_{i=1}^{5k}\varphi(\nu_i').$$

Therefore, for any $y \in Y$ with dimension $5k$ and its associated $\nu = m(y)$:

$$P^\nu(\nu) = F\left(\frac{1}{5k}\sum_{i=1}^{5k}\varphi(\nu_i)\right) \tag{2.17}$$

The same expression is valid for all $y \in Y$ with dimension n as the same reasoning can be applied between $n(y)$ and the least common multiple between $n(y)$ and 5.

Finally, transformations d' and G of respectively functions φ and F guarantee that the domain of image of d' is $[0, 1]$. Letting $d'(y_i, \overline{y}) = \frac{\varphi(d(y_i,\overline{y}))}{\varphi(1)}$ and $G(x) = F\big(x\varphi(1)\big)$, we have for all $y \in Y$:

$$P^\nu(\nu) = G\left(\frac{1}{n}\sum_{i=1}^{n} d'(y_i, \overline{y})\right) = P(y) \tag{2.18}$$

where G is a continuous and strictly increasing function and d' is a numerical representation of $\succeq$. As function G is strictly increasing, P is ordinally equivalent to $P' : Y \to [0, 1]$ with $P'(y) = \frac{1}{n}\sum_{i=1}^{n} d'(y_i, \overline{y})$. This proves P is an additive poverty index.

2.9.2 Proof of Lemma 6

The proof of Theorem 2 relies on Lemma 6, which gives a necessary condition and a sufficient condition for satisfying *Monotonicity in Income*. These conditions hold for the general domains of poverty lines, absolute-homothetic EOs and numerical representations.

The presentation of Lemma 6 requires introducing two definitions. For a given additive poverty index, the *degree of priority* of an income level over another at a certain mean income measures the ratio of the increase in the index if a marginal increase takes place at one income level rather than at the other.[48][49]

Definition 15 (Degree of Priority of y_i over y_j at $\overline{y}$). $DP_{ij}(\overline{y}) := \frac{\partial_1 d(y_i,\overline{y})}{\partial_1 d(y_j,\overline{y})}$

$DP_{ij}(\overline{y})$ can be interpreted as the priority given by the index to an income level y_i over another income level y_j when mean income is $\overline{y}$.

Monotonicity in Income sets a lower and an upper bound on the degrees of priority granted by additive indices. These bounds depend on the slopes of the equivalence curves at the bundles of the concerned agents. These slopes can be defined using the numerical representation.[50]

Definition 16 (Slope at $(y_i, \overline{y})$). $s(y_i, \overline{y}) := -\frac{\partial_2 d(y_i,\overline{y})}{\partial_1 d(y_i,\overline{y})}$

The two general conditions are the following.

Lemma 6 (Bounds on degrees of priority).
An additive poverty index based on an absolute-homothetic EO below an endogenous line satisfies Monotonicity in Income:

1. *(sufficient condition) if for all $\overline{y} > 0$ and $y_i, y_j < z(\overline{y})$, we have:*

$$s(y_j, \overline{y}) \leq DP_{ij}(\overline{y}) \tag{2.19}$$

2. *(necessary condition) only if for all $\overline{y} > 0$ with $z(\overline{y}) \leq \overline{y}$ and all $y_i, y_j < z(\overline{y})$, (2.19) holds.*

[48] The partial derivative of a function $f : X \to \mathbb{R}$ in the direction x_i at point $x \in X$ is denoted $\partial_i f(x)$.

[49] Numerical representations need not be differentiable everywhere. The definition of $DP_{ij}(\overline{y})$ at points for which d is not differentiable is given in Appendix 2.9.2 which treats non-differentiability of numerical representations.

[50] Again, the modification of this definition for points at which d is not differentiable is in Appendix 2.9.2. This definition allows attributing a unique value of the slope even at points for which the equivalence curves of the EO exhibit a kink.

Proof. Consider any additive index P based on an absolute-homothetic EO below an endogenous line. The index P satisfies *Monotonicity in Income* if and only if for all $y \in Y$ and $i \leq q$ we have $\partial_i P(y_1, \cdots, y_n) \leq 0$. By the additively separable form of P, this inequality becomes by chain derivation:[51]

$$\partial_1 d(y_i, \overline{y}) + \sum_{j=1}^{n} \partial_2 d(y_j, \overline{y}) \partial_i \overline{y} \leq 0. \tag{2.20}$$

From the definition of the mean, we have $\partial_i \overline{y} = \frac{1}{n}$. From the definition of $s(y_j, \overline{y})$, we get $\partial_2 d(y_j, \overline{y}) = -\partial_1 d(y_j, \overline{y}) s(y_j, \overline{y})$ for all $(y_j, \overline{y}) \in X$. Inequality (2.20) becomes:

$$\underbrace{\partial_1 d(y_i, \overline{y}) - \frac{1}{n} \sum_{j=1}^{n} \partial_1 d(y_j, \overline{y}) s(y_j, \overline{y})}_{L_{2.21}} \leq 0. \tag{2.21}$$

In the remainder of the proof, (2.21) is shown to imply the necessary and sufficient conditions linked to (2.19). Inequality (2.19) can be rewritten:

$$\underbrace{\partial_1 d(y_i, \overline{y}) - \partial_1 d(y_j, \overline{y}) s(y_j, \overline{y})}_{L_{2.22}} \leq 0. \tag{2.22}$$

Necessity of condition 2 is proved by contradiction. Assume (2.22) does not hold for some $y^1 \in Y$ with $z(\overline{y}^1) \leq \overline{y}^1$ and y_i^1, y_j^1 are such that $0 \leq y_i^1 < y_j^1 < z(\overline{y}^1)$. Therefore, at $(y_i^1, \overline{y}^1)$, $(y_j^1, \overline{y}^1) \in X_p$, we have for some $\ell > 0$ that $L_{2.22} = \ell$. I prove that for all $\epsilon > 0$, there exists $y^2 \in Y$ with $\overline{y}^2 = \overline{y}^1$ such that $|\ell - L_{2.21}(y^2)| < \epsilon$ and hence, for $\epsilon < \ell$, there exists an y^2 such that $L_{2.21}(y^2) > 0$, violating *Monotonicity in Income*. Construct y^2 such that

- $y_1^2 := y_i^1$,
- $y_k^2 := y_j^1$ for all k with $2 \leq k \leq n(y^2) - 1$ and
- $y_n^2 := n(y^2)\overline{y}^1 - \sum_{h=1}^{n(y^2)-1} y_h^2$.

Notice $y_n^2 \geq z(\overline{y}^1)$ since $\overline{y}^1 \geq z(\overline{y}^1)$, which implies $y^2 \in Y$. For distribution y^2, rembering that $\partial_1 d(y_n^2, \overline{y}^1) = 0$, we have:

$$\begin{aligned} \ell - L_{2.21}(y^2) &= L_{2.22} - L_{2.21}(y^2) \\ &= -\frac{1}{n(y^2)} \left(2\partial_1 d(y_j^1, \overline{y}^1) s(y_j^1, \overline{y}^1) - \partial_1 d(y_i^1, \overline{y}^1) s(y_i^1, \overline{y}^1) \right). \end{aligned}$$

In order to show that $|\ell - L_{2.21}(y^2)| < \epsilon$, two cases must be considered:

- Case 1: $\partial_1 d(y_j^1, \overline{y}^1)$ and $\partial_1 d(y_i^1, \overline{y}^1)$ are finite.
  The distance $|\ell - L_{2.21}(y^2)|$ can be made arbitrarly small by taking $n(y^2)$ sufficiently large, implying $L_{2.21}(y^2) > 0$, which violates (2.21) and hence *Monotonicity in Income*.

- Case 2: $\partial_1 d(y_j^1, \overline{y}^1)$ or $\partial_1 d(y_i^1, \overline{y}^1)$ are not finite.
  Observe first that if $\partial_1 d(y_i^1, \overline{y}^1) = -\infty$ and $\partial_1 d(y_j^1, \overline{y}^1)$ is finite, then (2.22)

[51] The case of points at which d is not differentiable is treated in Appendix 2.9.2.

must hold.
Assume $\partial_1 d(y_j^1, \overline{y}^1) = -\infty$. If (2.22) does not hold, then we have $s(y_j^1, \overline{y}^1) > 0$ as d is strictly decreasing in y_i. If $\partial_1 d(y_i^1, \overline{y}^1)$ is finite, then $L_{2.21}(y^2) > 0$ and *Monotonicity in Income* does not hold. If $\partial_1 d(y_i^1, \overline{y}^1) = -\infty$, by the continuity of d, there exists y_k^1 close to y_i^1 for which the equivalent of (2.22) does not hold and $\partial_1 d(y_k^1, \overline{y}^1)$ is finite, leading again to a violation of the axiom.

The case for which $0 \leq y_j^1 < y_i^1 < z(\overline{y}^1)$ leads to the same contradiction. The only difference lies in the construction of y^2: $y^2_{n(y^2)-1} := y_i^1$, $y_k^2 := y_j^1$ for all k with $1 \leq k \leq n(y^2) - 2$. The condition is therefore necessary.

Sufficiency of condition 1 follows from the fact that, if there exists an $y \in Y$ violating (2.21), inequality (2.22) is violated as well for a particular value of y_j. For all $y \in Y$ there exists $y_j^* \in \big[0, z(\overline{y})\big)$ such that, taking $y_j := y_j^*$ in $L_{2.22}$, we have $L_{2.21}(y) < L_{2.22}$, which is:

$$-\frac{1}{n}\sum_{j=1}^{n} \partial_1 d(y_j, \overline{y}) s(y_j, \overline{y}) < -\partial_1 d(y_j^*, \overline{y}) s(y_j^*, \overline{y}),$$

where the strict inequality comes from the presence of the non-poor agent n for whom $\partial_1 d(y_n, \overline{y}) = 0$. The key property for last inequality to hold is that $\partial_1 d(y_j, \overline{y})$ and $s(y_j, \overline{y})$ depend on the income of other agents only through their impact on mean income $\overline{y}$. At mean income $\overline{y}$, y_j^* is obtained by solving the following problem:

$$y_j^* := \arg \max_{y_j \in [0, z(\overline{y}))} -\partial_1 d(y_j, \overline{y}) s(y_j, \overline{y}).$$

■

The symmetry of degrees of priority implies that the lower bound given in (2.19) is associated with an upper bound.[52] Lemma 6 shows that the steeper the equivalence curves, the narrower is the range of acceptable degrees of priority. These rather obscure constraints have strong implications that are best illustrated on specific domains of poverty lines, EO's and families of numerical representations.

Non-differentiability of numerical representation d

I extend in this subsection the definitions of degrees of priorities and slopes for bundles at which the numerical representation is not differentiable. I show how these extended definitions allows Lemma 6 to hold even at those bundles and hence everywhere for absolute-homothetic EOs.

Function d is differentiable almost everywhere as the function d is continuous. Consider any $(y_i^1, \overline{y}^1)$, $(y_j^1, \overline{y}^1) \in X_p$ at which d is not differentible. The definition of

[52] The symmetric definition of $DP_{ij}(\overline{y})$ implies that $DP_{ij}(\overline{y}) = \frac{1}{DP_{ji}(\overline{y})}$, at least at bundles at which d is differentiable. If $s(y_i, \overline{y}) \geq 0$, inequality (2.19) could be completed by a second inequality it implies, which gives the associated upper bound: $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y}) \leq \frac{1}{s(y_i, \overline{y})}$.

DP_{ij} at these bundles is given by

$$DP_{ij}(\overline{y}^1) := \frac{\lim_{y_i \to y_i^{1+}} \partial_1 d(y_i, \overline{y}^1)}{\lim_{y_j \to y_j^{1-}} \partial_1 d(y_j, \overline{y}^1)}.$$

Either these limits take non-negative finite values or they tend to infinity, showing that $DP_{ij} \in [0, \infty)$.[53] For any $(y_j^1, \overline{y}^1) \in X_p$ at which d is not differentible, the definition of the slope becomes

$$s(y_j^1, \overline{y}^1) := -\frac{\lim_{\overline{y} \to \overline{y}^{1+}} \partial_2 d(y_j^1, \overline{y})}{\lim_{y_j \to y_j^{1-}} \partial_1 d(y_j, \overline{y}^1)} = \begin{cases} 0 & \text{if } y_j \leq z^a, \\ \frac{y_j - z^a}{z(\overline{y}) - z^a} \partial z(\overline{y}) & \text{otherwise.} \end{cases}$$

The definition of the slope at bundle $(y_j^1, \overline{y}^1)$ where d is not differentiable implies

$$\lim_{\overline{y} \to \overline{y}^{1+}} \partial_2 d(y_j^1, \overline{y}) = - \lim_{y_j \to y_j^{1-}} \partial_1 d(y_j, \overline{y}^1) s(y_j^1, \overline{y}^1).$$

From the previous equation, we can extend (2.22) in Lemma 6, which must now be compared with an extended version of (2.20), that is obtained by chain derivation of P at y:

$$\lim_{y_j \to y_i^{1+}} \partial_1 d(y_i, \overline{y}) + \sum_{j=1}^{n} \lim_{\overline{y} \to \overline{y}^{1+}} \partial_2 d(y_j, \overline{y}) \partial_i \overline{y} \leq 0.$$

The reasoning given in the proof of Lemma 6 is then valid even at those points. This extension of the validity of Lemma 6 is only necessary for the proof of Theorem 4. Indeed, other theorems relies on families of numerical representations that are differentiable everywhere.

Observe that non-smooth equivalence curves are not ruled out from absolute-homothetic EOs. Indeed, poverty lines can exhibit kinks, as it is the case at y^k for piecewise-linear lines. This non-smoothness is not problematic as the extended definition of slope given above guarantees there is a unique value of slope at these bundles. The evolution of slopes with $\overline{y}$ is not continuous at $\overline{y}^k$, but this does not affect Lemma 6, which provides conditions to be checked independently at each particular value of mean income $\overline{y}$.

2.9.3 Proof of Theorem 2

Take any monotonic endogenous line z and any absolute-homothetic EO $\succeq$ below z. Take additive index P whose numerical representation d of $\succeq$ belongs to the extended FGT family. I prove Theorem 2 claim by claim.

Claim 1: P satisfies *Monotonicity in Income* only if $\alpha = 1$.

[53] It can be that for $y_i \neq y_j$ we have $DP_{ij} = \frac{\infty}{\infty}$, meaning DP_{ij} is not well-defined. I show that ignoring these cases in the necessary and sufficient condition is not problematic. Two cases can happen. Case 1: we have $y_i \leq y_j \leq z^a$ and hence $s(y_i, \overline{y}) = s(y_j, \overline{y}) = 0$. This case leads to no violation of inequality (2.19) as $DP_{ij} \in [0, \infty)$. Case 2: we have either $y_i > z^a$ or $y_j > z^a$ (assume the later without loss of generality). Then there always exists $y_k < z(\overline{y})$ such that $DP_{kj} = 0$ and inequality (2.19) is violated as $s(y_j, \overline{y}) > 0$.

The numerical representation of P belongs to the extended FGT family which means there exists $\overline{y}^r \geq 0$ such that for all $(y_i, \overline{y}) \in X_p$ we have

$$d(y_i, \overline{y}) = \left(\frac{z(\overline{y}^r) - e^r(y_i, \overline{y})}{z(\overline{y}^r)}\right)^{\alpha}. \quad (2.23)$$

I show that if $\alpha \neq 1$ then the necessary condition given in Lemma 6 (see Appendix 2.9.2) for the associated additive poverty index P to satisfy *Monotonicity in Income* is violated.

Since line z is monotonic, there exists $g > 0$ with $g \geq z(g)$ such that $s(g) > 0$. Consider any $(y_i^1, \overline{y}^1) \in X_p$ with $\overline{y}^1 = g$ and $y_i^1 > z^a$. By the monotonicity of z, we have hence that $\overline{y}^1 \geq z(\overline{y}^1)$. From the necessary condition in Lemma 6, if there exists y_j^1 with $y_j^1 < z(\overline{y}^1)$ and $s(y_j^1, \overline{y}^1) > DP_{ij}(\overline{y}^1)$ then *Monotonicity in Income* does not hold.[54] I show below there exists y_j^1 with $y_i^1 < y_j^1 < z(\overline{y}^1)$ leading to a violation of the necessary condition when $\alpha \neq 1$.

The degree of priority given by P to agent i over j, when $y_i^1 \leq y_j^1$ is obtained by chain derivation of (2.23):

$$DP_{ij}(\overline{y}^1) = \frac{\partial_1 d(y_i^1, \overline{y}^1)}{\partial_1 d(y_j^1, \overline{y}^1)} = \left(\frac{z(\overline{y}^r) - e^r(y_i^1, \overline{y}^1)}{z(\overline{y}^r) - e^r(y_j^1, \overline{y}^1)}\right)^{\alpha-1} \underbrace{\frac{\partial_1 e^r(y_i^1, \overline{y}^1)}{\partial_1 e^r(y_j^1, \overline{y}^1)}}_{F1}, \quad (2.24)$$

where $\overline{y}^r$ denotes the value of mean income at which d takes the exponential mathematical expression. Factor F1 in (2.24) is equal to one because the EO is absolute-homothetic. Indeed, absolute-homotheticity implies that for all $\overline{y} > 0$ and $y_i, y_j \in [z^a, z(\overline{y})]$ we have

$$\frac{e^r(y_j, \overline{y}) - z^a}{e^r(y_i, \overline{y}) - z^a} = \frac{y_j - z^a}{y_i - z^a}. \quad (2.25)$$

Therefore (2.24) can be simplified to

$$DP_{ij}(\overline{y}^1) = \left(\frac{z(\overline{y}^r) - e^r(y_i^1, \overline{y}^1)}{z(\overline{y}^r) - e^r(y_j^1, \overline{y}^1)}\right)^{\alpha-1}. \quad (2.26)$$

I now prove that the necessary condition is violated for y_j^1 sufficiently close to $z(\overline{y}^1)$. Three cases must be considered depending on the value taken by α.

- Case 1: $0 < \alpha < 1$:
  When y_j^1 tends to $z(\overline{y}^1)$, we have that $e^r(y_j, \overline{y})$ tends to $z(\overline{y}^r)$. From the exponential functional form of $DP_{ij}(\overline{y}^1)$, for all $\epsilon > 0$, there exists hence $y_j^1 \in \left[y_i^1, z(\overline{y}^1)\right)$ such that
  $$DP_{ij}(\overline{y}^1) = \left(\frac{z(\overline{y}^r) - e^r(y_i, \overline{y})}{z(\overline{y}^r) - e^r(y_j, \overline{y})}\right)^{\alpha-1} < \epsilon.$$
  As the poverty line is monotonic and $\overline{y}^1 = g$, we have $s(z(\overline{y}^1)) > 0$. As the EO is absolute-homothetic and $y_i^1 > z^a$, we have $s(y_i^1, \overline{y}^1) > 0$ and for all $y_j^1 > y_i^1$ we have $s(y_j^1, \overline{y}^1) > s(y_i^1, \overline{y}^1)$. As a result, for any $\epsilon < s(y_i^1, \overline{y}^1)$ we have $s(y_j^1, \overline{y}^1) > DP_{ij}(\overline{y}^1)$ and the necessary condition is violated.

[54] See Appendix 2.9.2 for the definitions of slope $s(y_j^1, \overline{y}^1)$ and degree of priority $DP_{ij}(\overline{y}^1)$.

- Case 2: $\alpha > 1$:
  As the numerical representation is differentiable for all y_i^1 and y_j^1 with $z^a < y_i^1 < y_j^1$, we have $DP_{ji}(\overline{y}^1) = \frac{1}{DP_{ij}(\overline{y}^1)}$. From the reasoning given for the case $0 < \alpha < 1$, we have that for all $\epsilon > 0$, there exists $y_j^1 \in \big[y_i^1, z(\overline{y}^1)\big)$ such that $DP_{ji}(\overline{y}^1) < \epsilon$. This leads to a violation of *Monotonicity in Income* for identic reasons.

- Case 3: $\alpha = 0$:
  Index P is an increasing transformation of the Head-Count Ratio. *Monotonicity in Income* is violated for any $y \in Y$ with $\overline{y} = g$ and one non-poor agent i has income $y_i = z(\overline{y})$.

Claim 2: P satisfies *Monotonicity in Income* and *Transfer among Poor* if and only if $\alpha = 1$ and $\overline{y}^r = 0$.

By Claim 1, P satisfies *Monotonicity in Income* only if $\alpha = 1$. Claim 2 is therefore proven by the combination of steps 1 and 2.

Step 1: If $\alpha = 1$, then P satisfies *Transfer among Poor* if and only if $\overline{y}^r = 0$.

P satisfies *Transfer among Poor* if and only if the numerical representation d is convex at all values of mean income. Formally, P satisfies *Transfer among Poor* if and only if for all $\overline{y} > 0$ and all $y_i, y_j \in [0, z(\overline{y}))$ with $y_i < y_j$ we have $DP_{ij}(\overline{y}) \geq 1$. Lemma 7 shows that we have $DP_{ij}(\overline{y}) \neq 1$ only in the case $y_i < z^a < y_j$.

Lemma 7. *Let $\succeq$ be an absolute-homothetic EO below an endogenous line. Let d be a numerical representation of $\succeq$ in the extended FGT family with $\alpha = 1$. For all $(y_i, \overline{y}), (y_j, \overline{y}) \in X_p$ with $y_i \leq y_j$, if $DP_{ij}(\overline{y}) \neq 1$, then $y_i < z^a < y_j$ and*

$$DP_{ij}(\overline{y}) = \frac{z(\overline{y}) - z^a}{z(\overline{y}^r) - z^a}. \tag{2.27}$$

Proof. Consider any $(y_i, \overline{y}), (y_j, \overline{y}) \in X_p$ with $y_i \leq y_j$. Given $\alpha = 1$, the value taken by $DP_{ij}(\overline{y})$ depends only on $\overline{y}$ and on the relative positions of y_i, y_j and z^a. Four cases must be considered.

- Case 1: $y_i = y_j$
  $DP_{ij}(\overline{y}) = 1$ by the definition of $DP_{ij}(\overline{y})$.

- Case 2: $y_i < y_j \leq z^a$
  Equation (2.24) holds as it does not depend on the particular value of mean income $\overline{y}$. By absolute-homotheticity we have for all $\overline{y} > 0$ and $y_i \leq z^a$ that $e^r(y_i, \overline{y}) = y_i$. As a result (2.26) holds as well. Replacing $\alpha = 1$ leads to $DP_{ij}(\overline{y}) = 1$.

- Case 3: $z^a \leq y_i < y_j$
  Equation (2.26) holds. Replacing $\alpha = 1$ leads to $DP_{ij}(\overline{y}) = 1$.

- Case 4: $y_i < z^a < y_j$
  As $\alpha = 1$, the numerical representation at any point $(y_i, \overline{y}) \in X_p$ is
  $$d(y_i, \overline{y}) = \left(\frac{z(\overline{y}) - e^r(y_i, \overline{y})}{z(\overline{y})} \right).$$
  As the EO is absolute-homothetic, for any $y_i \leq z^a$ we have $e^r(y_i, \overline{y}) = y_i$. As a result we have for any $y_i \leq z^a$ that
  $$\partial_1 d(y_i, \overline{y}) = \frac{-1}{z(\overline{y}^r)}.$$
  As the EO is absolute-homothetic, for any $y_j \geq z^a$ we have that
  $$e^r(y_j, \overline{y}) - z^a = (y_j - z^a) \frac{z(\overline{y}^r) - z^a}{z(\overline{y}) - z^a},$$
  implying for any $y_j \geq z^a$ that
  $$\partial_1 d(y_j, \overline{y}) = \frac{-1}{z(\overline{y}^r)} \left(\frac{z(\overline{y}^r) - z^a}{z(\overline{y}) - z^a} \right).$$
  By the definition of $DP_{ij}(\overline{y})$, we find for any $y_i < z^a < y_j$ that (2.27) holds.

■

I show that $\overline{y}^r = 0$ is sufficient and necessary for P to satisfy *Transfer among Poor*.

- Case $\overline{y}^r = 0$ (sufficiency)
  This case is such that $\overline{y}^r < \overline{y}$ for all $\overline{y} > 0$. As z is assumed monotonic, we have that $z(\overline{y}^r) \leq z(\overline{y})$. Then for all $y_i, y_j \in [0, z(\overline{y}))$ with $y_i < z^a < y_j$ we have by (2.27) and $z(\overline{y}^r) \leq z(\overline{y})$ that $DP_{ij}(\overline{y}) \geq 1$. This implies by Lemma 7 that, when $y_i \leq y_j$, we have $DP_{ij}(\overline{y}) \geq 1$ and hence the sufficient condition for *Transfer among Poor* holds.

- Case $\overline{y}^r > 0$ (necessity)
  This case is such that there exists $\overline{y} < \overline{y}^r$ such that $z(\overline{y}) < z(\overline{y}^r)$. Indeed, if the poverty line is flat for all $\overline{y} < \overline{y}^r$, then the numerical representation is linear in $\overline{y} = 0$ as the EO is absolute-homothetic. Therefore the numerical representation is equivalent to $\overline{y}^r = 0$ and we have $\overline{y}^r = 0$.

  At $\overline{y} < \overline{y}^r$ such that $z(\overline{y}) < z(\overline{y}^r)$ we have for any $y_i < z^a < y_j$ that $DP_{ij}(\overline{y}) < 1$ from (2.27), violating the necessary condition for *Transfer among Poor*.

Step 2: If $\alpha = 1$ and $\overline{y}^r = 0$, then P satisfies *Monotonicity in Income*.

The sufficient condition for *Monotonicity in Income* given in Lemma 6 requires that for all $\overline{y} > 0$ and $y_i, y_j < z(\overline{y})$, we have $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$.

As the EO satisfies Translation Monotonicity, we have for all $(y_i, \overline{y}) \in X_p$ that $s(y_j, \overline{y}) \leq 1$. The sufficient condition can therefore only be violated if $DP_{ij}(\overline{y}) < 1$.

From Lemma 7, the case $DP_{ij}(\overline{y}) \neq 1$ can only happen if one agent is absolutely poor (income below z^a) and the other relatively poor (income above z^a). As $\overline{y}^r =$

0, $\alpha = 1$ and since the line is monotonic, (2.27) shows that the relatively poor agent cannot have a priority over the absolutely poor agent strictly larger than 1. Therefore, the case $DP_{ij}(\overline{y}) < 1$ only happens if $y_j < z^a < y_i$.

As the EO is absolute-homothetic, if $y_j < z^a$ then $s(y_i, \overline{y}) = 0$ and the sufficient condition holds since $DP_{ij}(\overline{y})$ is non-negative by definition.

Piecewise-linear poverty line

This subsection shows that if the endogenous line is piecewise-linear, then there exists an upper-bound for the value of reference mean income below which the PGR at $\overline{y}^r$ satisfies *Monotonicity in Income*.

Theorem 8 (Upper-bound for reference mean income)**.**
Let z be a piecewise-linear poverty line with $\overline{y}^k \geq z^0$ and slope $\overline{s} > 0$. Let P be an additive poverty index based on an absolute-homothetic EO below z with a numerical representation in the extended FGT family with $\alpha = 1$.

1. *P satisfies Monotonicity in Income if and only if:*
$$\overline{y}^r \leq \overline{y}^k + \left(\frac{1-\overline{s}}{\overline{s}^2}\right)(z^0 - z^a).$$

Proof.
Step 1: P satisfies *Monotonicity in Income* if and only if for all $\overline{y} \geq \overline{y}^k$ and all $y_i, y_j < z(\overline{y})$, we have $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$.

Given z is piecewise-linear, for all $\overline{y} \leq \overline{y}^k$ and all $y_i < z(\overline{y})$ we have $s(y_i, \overline{y}) = 0$. As a result, for all $\overline{y} \leq \overline{y}^k$ and all $y_i, y_j < z(\overline{y})$ inequality $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$ holds. Therefore, the necessary condition for *Monotonicity in Income* given in Lemma 6 is also sufficient.

Step 2: P satisfies *Monotonicity in Income* if and only if for all $\overline{y} \geq \overline{y}^k$ and $y_j < z(\overline{y})$ we have $s(y_j, \overline{y}) \leq \frac{z(\overline{y})-z^a}{z(\overline{y}^r)-z^a}$.

As the EO satisfies Translation Monotonicity we have $s(y_j, \overline{y}) \leq 1$ for all $(y_j, \overline{y}) \in X_p$. Assuming without loss of generality that $y_i \leq y_j < z(\overline{y})$, by Lemma 7 we have that inequality $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$ is violated only if $y_i < z^a < y_j$. In that case, by (2.27) we get
$$DP_{ij}(\overline{y}) = \frac{z(\overline{y}) - z^a}{z(\overline{y}^r) - z^a}.$$
Therefore, condition of Step 2 is a simplified version of the necessary and sufficient condition of Step 1.

Step 3: P satisfies *Monotonicity in Income* if and only if
$$\overline{y}^r \leq \overline{y}^k + \left(\frac{1-\overline{s}}{\overline{s}^2}\right)(z^0 - z^a).$$

For all $\overline{y} \geq \overline{y}^k$, slope $s(y_j, \overline{y})$ is maximal and tends to $\bar{s}$ when y_j tends to $z(\overline{y})$. When $y_i < z^a < y_j$, considering any $y_j \geq z^a$ does not affect the value of $DP_{ij}(\overline{y})$ found in Step 2. Therefore replacing $s(y_j, \overline{y})$ by $\bar{s}$ in the condition of Step 2 is without loss of generality.

Given $\overline{y}^r$, $DP_{ij}(\overline{y})$ is weakly decreasing in $\overline{y}$ (constant on $\overline{y} \leq \overline{y}^k$) and reach a minimal value for $\overline{y} = \overline{y}^k$. Therefore, if the inequality given in Step 2 holds for $\overline{y} = \overline{y}^k$, then it holds for all $\overline{y} > 0$. Therefore *Monotonicity in Income* holds if and only if:

$$\bar{s} \leq \frac{z(\overline{y}^k) - z^a}{z(\overline{y}^r) - z^a} = \frac{z^0 - z^a}{z(\overline{y}^r) - z^a},$$

which yields the desired threshold for $\overline{y}^r$ as for all $\overline{y}^r \geq \overline{y}^k$ we have $z(\overline{y}^r) = z^0 + \bar{s}(\overline{y}^r - \overline{y}^k)$. ■

2.9.4 Proof of Theorem 3

Take any linear line z with $\bar{s} > 0$ and any additive index P with a numerical representation d of the homothetic EO below z belonging to the quadratic family.[55] This proof is made in two steps, which together constitute the proof.

STEP 1: P satisfies *Monotonicity in Income* if and only if for some arbitrary $\overline{y}^1 > 0$ and all $y_i, y_j \in [0, z(\overline{y}^1))$ we have $s(y_j, \overline{y}^1) \leq DP_{ij}(\overline{y}^1)$.

Take any $\overline{y}^1$ with $z(\overline{y}^1) \leq \overline{y}^1$. As the poverty line is linear and hence $\bar{s} > 0$, such $\overline{y}^1$ exists. As shown in the necessary condition of Lemma 6, P satisfies *Monotonicity in Income* only if for all $y_i, y_j \in [0, z(\overline{y}^1))$ we have $s(y_j, \overline{y}^1) \leq DP_{ij}(\overline{y}^1)$.

By assumption the EO is homothetic. I show that homotheticity implies that the degree of priority of one equivalence level over another does not depend on mean income. For all $\overline{y}^1, \overline{y}^2 > 0$, $(y_i, \overline{y}^1), (y_j, \overline{y}^1) \in X_p$ if $y_k := e^2(y_i, \overline{y}^1)$ and $y_\ell := e^2(y_j, \overline{y}^1)$ then $DP_{ij}(\overline{y}^1) = DP_{k\ell}(\overline{y}^2)$. Homotheticity implies that

$$e^2(y_j, \overline{y}^1) = \frac{y_j}{y_i} e^2(y_i, \overline{y}^1).$$

By chain derivation, the degree of priority of y_i over y_j at $\overline{y}^1$ is hence:

$$DP_{ij}(\overline{y}^1) = \frac{\partial_1 d\big(e^2(y_i, \overline{y}^1), \overline{y}^2\big)}{\partial_1 d\big(e^2(y_j, \overline{y}^1), \overline{y}^2\big)} \frac{\partial_1 e^2(y_i, \overline{y}^1)}{\partial_1 e^2(y_j, \overline{y}^1)} = DP_{k\ell}(\overline{y}^2) \frac{\frac{e^2(y_j, \overline{y}^1)}{y_j}}{\frac{e^2(y_i, \overline{y}^1)}{y_i}} = DP_{k\ell}(\overline{y}^2).$$

By assumption, the poverty line z is linear which, together with a monotonic EO implies that all bundles yielding the same equivalence level have a constant slope, for all values of mean income. For all $(y_i, \overline{y}^1) \in X_p$ and all $(y_k, \overline{y}^2) \in X_p$ with $(y_k, \overline{y}^2) \sim (y_i, \overline{y}^1)$, we have:

$$s(y_i, \overline{y}^1) = \bar{s} \frac{y_i}{z(\overline{y}^1)} = \bar{s} \frac{y_k}{z(\overline{y}^2)} = s(y_k, \overline{y}^2). \tag{2.28}$$

[55] If $\bar{s} = 0$, then the line is absolute and P satisfies *Monotonicity in Income* since in that case *Monotonicity in Income* is implied by *Domination among Poor*.

Therefore, if for all $y_i, y_j \in [0, z(\overline{y}^1))$ we have $s(y_j, \overline{y}^1) \leq DP_{ij}(\overline{y}^1)$, then for all $\overline{y} > 0$ and all $y_i, y_j \in [0, z(\overline{y}))$ we have $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$. Therefore the sufficient condition in Lemma 6 holds as well in that case.

STEP 2: For any $\overline{y} > 0$, we have $s(y_j, \overline{y}^1) \leq DP_{ij}(\overline{y}^1)$ for all $y_i, y_j \in [0, z(\overline{y}))$ if and only if inequalities (2.9) hold.

Take any $\overline{y} > 0$. For all $y_i \in [0, z(\overline{y}))$, since d belongs to the quadratic family:

$$\partial_1 d(y_i, \overline{y}) = -\frac{1}{z(\overline{y})}\Big(1 + \alpha\Big(1 - 2\frac{y_i}{z(\overline{y})}\Big)\Big).$$

Therefore for all $y_i, y_j \in [0, z(\overline{y}))$, using the expression of $s(y_i, \overline{y})$ given in 2.28, inequality $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$ is rewritten:

$$\underbrace{\frac{y_j}{z(\overline{y})} \frac{1 + \alpha\Big(1 - 2\frac{y_j}{z(\overline{y})}\Big)}{1 + \alpha\Big(1 - 2\frac{y_i}{z(\overline{y})}\Big)}}_{L_{2.29}} \leq \frac{1}{\bar{s}}. \tag{2.29}$$

Two cases must be considered for this inequality:

- Case 1: $\alpha < 0$.
  $L_{2.29}$ is maximal when (i) $y_i = 0$ and (ii) y_j tends to $z(\overline{y})$, implying that $s(y_j, \overline{y})$ tends to $\bar{s}$. Replacing those values yields the lower bound on α.

- Case 2: $\alpha \geq 0$.
  $L_{2.29}$ is maximal when (i) y_i tends to $z(\overline{y})$ and (ii) $\frac{y_j}{z(\overline{y})} = \frac{(1+\alpha)}{4\alpha}$ for all α with $\frac{1}{3} \leq \alpha \leq 1$. Replacing those values yields the upper bound on α. For all $\alpha \in \big[0, \frac{1}{3}\big]$, inequality (2.29) is respected for all $y_i, y_j \in [0, z(\overline{y}))$ as $\bar{s} \leq 1$.

2.9.5 Proof of Theorem 4

Take any piecewise-linear line z with $\overline{y}^k \geq z^0$ and $\bar{s} > 0$. Take any $x^* > 0$ with $x^* < z^0$. Let P be an additive index based on an EO in $\mathcal{R}^{HH}(z, x^*)$. Let $\succeq_{s_x}$ be an EO belonging to the subdomain $\mathcal{R}^{HH}(z, x^*)$ and hence $x^0 = x^*$.

First, I prove Claim 1. If $s_x = 0$, then $\succeq_{s_x}$ is absolute-homothetic and Theorem 2 shows that if the numerical representation of P is the PGR at the origin, then it satisfies both properties. I focus hence on proving the claim for EOs with $s_x > 0$. The proof relies on Lemma 8 giving a necessary and sufficient condition for an index to satisfy both properties.

Lemma 8. *P satisfies Monotonicity in Income and Transfer among Poor if and only if for all $\overline{y} \geq \overline{y}^k$ and all $y_i, y_j \in [0, z(\overline{y}))$ with $y_i < y_j$ we have $1 \leq DP_{ij}(\overline{y}) \leq \frac{1}{s(y_i, \overline{y})}$.*

Proof.
P satisfies *Transfer among Poor* if and only if for all $\overline{y} > 0$ and all $y_i, y_j \in [0, z(\overline{y}))$ with $y_i \leq y_j$ we have $1 \leq DP_{ij}(\overline{y})$. Given the poverty line is piecewise-linear and the EO is homothetic-homothetic, for all $\overline{y}, \overline{y}'$ with $\overline{y} < \overline{y}' \leq \overline{y}^k$ and all $y_i, y_j \in [0, z(\overline{y}))$

we have $DP_{ij}(\overline{y}) = DP_{ij}(\overline{y}')$. This implies that if the condition for *Transfer among Poor* holds for all $\overline{y} \geq \overline{y}^k$, then it holds for all $\overline{y} > 0$.

Given the poverty line is piecewise-linear and the EO is homothetic-homothetic, for all $\overline{y} < \overline{y}^k$ and all $y_i, y_j \in [0, z(\overline{y}))$, inequality $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$ is trivially satisfied as $s(y_j, \overline{y}) = 0$. By assumption we have $\overline{y}^k \geq z^0$, which implies $z(\overline{y}^k) \leq \overline{y}^k$ and hence $z(\overline{y}) \leq \overline{y}$ for all $\overline{y} \geq \overline{y}^k$. The necessary condition for *Monotonicity in Income* stated in Lemma 6 is therefore also sufficient: P satisfies *Monotonicity in Income* if and only if for all $\overline{y} \geq \overline{y}^k$ and all $y_i, y_j \in [0, z(\overline{y}))$ we have $s(y_j, \overline{y}) \leq DP_{ij}(\overline{y})$.[56] If P satisfies *Transfer among Poor*, this condition is met for all $y_i \leq y_j$ as $s(y_j, \overline{y}) \leq 1$ by Translation Monotonicity. For all $y_j < y_i$, the condition for *Monotonicity in Income* is based on inequality $s(y_i, \overline{y}) \leq DP_{ji}(\overline{y})$. As $DP_{ij}(\overline{y}) = \frac{1}{DP_{ji}(\overline{y})}$, we have that P satisfies both *Monotonicity in Income* and *Transfer among Poor* if and only if for all $\overline{y} \geq \overline{y}^k$ and all $y_i, y_j \in [0, z(\overline{y}))$ with $y_i < y_j$ we have $1 \leq DP_{ij}(\overline{y}) \leq \frac{1}{s(y_i, \overline{y})}$. ■

Given a particular EO, choosing an additive index P is equivalent to choosing its numerical representation d. At the reference mean income $\overline{y}^r$ at which d is expressed, selecting d is equivalent to select for all $y_i, y_j \in [0, z(\overline{y}^r))$ with $y_i < y_j$ the degrees of priority $DP_{ij}(\overline{y}^r)$.[57] Each income level y_i at $\overline{y}^r$ is associated to an equivalence level corresponding to the equivalence curve of the EO passing through the bundle $(y_i, \overline{y}^r)$. Being selected at $\overline{y}^r$, the degrees of priority between two *equivalence levels* evolve with mean income $\overline{y}$, according to the evolution of the equivalence curves of the EO. I compute below how those degrees of priority between two equivalence levels evolve with mean income $\overline{y}$. Then, I derive the conditions on the slope s_x of the homothetic-homothetic EO under which it is possible that, for the whole range of mean incomes $\left[\overline{y}^k, \infty\right)$, the degrees of priority stay inside $\left[1, \frac{1}{s(y_i, \overline{y})}\right]$. The proof of Claim 1 is in three steps.

STEP 1: Evolution of DP_{ij} with mean income depends on s_x.

Take $\overline{y}^k$ as reference mean income. Any reference mean income is taken without loss of generality as there is no constraint on the mathematical expression of d.

Consider any two $y_i, y_j \in [0, z(\overline{y}^k))$ with $y_i < y_j$. Let $e^k : X_p \to [0, z(\overline{y}^k)]$ be the equivalent income function at $\overline{y}^k$. For any $\overline{y} > \overline{y}^k$, consider the bundles $(y_i', \overline{y})$ and $(y_j', \overline{y})$ yielding equivalent individual poverty, that is $y_i = e^k(y_i', \overline{y})$ and $y_j = e^k(y_j', \overline{y})$. As EO $\succeq_{s_x}$ is homothetic-homothetic, we have for all $\overline{y} \geq \overline{y}^k$ and all $y_\ell \in \left[0, x(\overline{y})\right)$ that

$$e^k(y_\ell, \overline{y}) = y_\ell \frac{x^0}{x(\overline{y})},$$

[56] Although Lemma 6 is only proven for absolute-homothetic EOs, extending its validity to homothetic-homothetic EOs is staithforward.

[57] This selection is under the constraint that for any $y_\ell \in [0, z(\overline{y}^r))$ with $y_i < y_\ell < y_j$ we have $DP_{ij}(\overline{y}^r) = DP_{i\ell}(\overline{y}^r) DP_{\ell j}(\overline{y}^r)$, at least if d is differentiable at $(y_\ell, \overline{y}^r)$.

and for all $y_\ell \in \big[x(\overline{y}), z(\overline{y})\big)$ that

$$e^k(y_\ell, \overline{y}) = x^0 + \frac{z^0 - x^0}{z(\overline{y}) - x(\overline{y})}\big(y_\ell - x(\overline{y})\big).$$

The evolution of DP_{ij} as a function of $\overline{y}$ depends on the relative positions of y_i, y_j and $x(\overline{y}^k) = x^0$. Three cases must be considered.

- Case 1: $y_i < y_j \leq x^0$.
  A direct extension of the reasoning on homotheticity in Step 1 of Theorem 3's proof shows that for all $\overline{y} > \overline{y}^k$ and all $y_i, y_j < x(\overline{y}^k)$ we have $DP_{ij}(\overline{y}) = DP_{ij}(\overline{y}^k)$. This implies that if the necessary and sufficient condition of Lemma 8 is met at $\overline{y}^k$, it is met for all $\overline{y} \geq \overline{y}^k$.

- Case 2: $x^0 \leq y_i < y_j$.
  We have $DP_{ij}(\overline{y}) = DP_{ij}(\overline{y}^k)$. Again, checking the condition at $\overline{y}^k$ is necessary and sufficient.

- Case 3: $y_i < x^0 < y_j$.
  By chain derivation, we obtain successively:

$$\begin{aligned} DP_{ij}(\overline{y}) &= \frac{\partial_1 d\big(e^k(y_j,\overline{y}),\overline{y}^k\big)}{\partial_1 d\big(e^k(y_j,\overline{y}),\overline{y}^k\big)}\frac{\partial_1 e^k(y_i,\overline{y})}{\partial_1 e^k(y_j,\overline{y})} = DP_{ij}(\overline{y}^k)\left(\frac{x^0}{x(\overline{y})}\frac{z(\overline{y})-x(\overline{y})}{z^0-x^0}\right) \\ &= DP_{ij}(\overline{y}^k)\left(\frac{x^0}{x^0+s_x(\overline{y}-\overline{y}^k)}\frac{(z^0-x^0)+(s_z-s_x)(\overline{y}-\overline{y}^k)}{z^0-x^0}\right), \end{aligned}$$

  and finally

$$DP_{ij}(\overline{y}) = DP_{ij}(\overline{y}^k)\frac{x^0}{z^0-x^0}\left(\frac{(z^0-x^0)+(s_z-s_x)(\overline{y}-\overline{y}^k)}{x^0+s_x(\overline{y}-\overline{y}^k)}\right). \tag{2.30}$$

  Taking the partial derivative of $DP_{ij}(\overline{y})$ with respect to mean income yields

$$\frac{\partial DP_{ij}(\overline{y})}{\partial \overline{y}} = DP_{ij}(\overline{y}^k)\frac{x^0}{z^0-x^0}\frac{s_z x^0 - s_x z^0}{\big(x^0+s_x(\overline{y}-\overline{y}^k)\big)^2}.$$

  There are three subcases to consider, depending on the value of s_x.

  - Subcase 3.1: $s_x \in \left(0, \frac{x^0}{z^0}s_z\right)$.
    The partial derivative of $DP_{ij}(\overline{y})$ is strictly increasing for all $\overline{y} > \overline{y}^k$.
  - Subcase 3.2: $s_x \in \left(\frac{x^0}{z^0}s_z, s^z\right]$.
    The partial derivative of $DP_{ij}(\overline{y})$ is strictly decreasing for all $\overline{y} > \overline{y}^k$.
  - Subcase 3.3: $s_x = \frac{x^0}{z^0}s_z$.
    The partial derivative of $DP_{ij}(\overline{y})$ is zero for all $\overline{y} > \overline{y}^k$. For $s_x = s_x^h$, the EO $\succeq_{s_x}$ corresponds to an homothetic EO. If the numerical representation is the PGR at the origin, then P respects both properties as shown in Theorem 2.

The degree of priority between two equivalence levels evolve with $\overline{y}$ only if $y_i < x^0 < y_j$. As shown by the three subcases, this evolution depends on the value taken by s_x. I study in Step 2 the conditions under which an EO in subcase 3.1 admits a numerical representation such that P satisfies both properties. In Step 3, I study those conditions in subcase 3.2.

STEP 2: For subcase 3.1, derive the lower bound $\underline{s}_x$ for s_x above which an additive index satisfies the necessary and sufficient conditions of Lemma 8.

Subcase 3.1 is such that $0 < s_x < \frac{x^0}{z^0} s_z$. Step 1 showed for these values of s_x that $DP_{ij}(\overline{y})$ is strictly increasing in $\overline{y}$. The necessary and sufficient condition for both properties given in Lemma 8 requires that we have $1 \leq DP_{ij}(\overline{y}) \leq \frac{1}{s(y_i,\overline{y})}$ for all $\overline{y} \geq \overline{y}^k$. As $DP_{ij}(\overline{y})$ is strictly increasing with $\overline{y}$, it is sufficient to check these inequalities at the boundaries of the domain for mean income, that is at $\overline{y} = \overline{y}^k$ and when $\overline{y}$ tends to ∞. From (2.30), the condition in Lemma 8 is satisfied only if for all $y_i, y_j \in [0, z(\overline{y}^k))$ with $y_i < x(\overline{y}^k) < y_j$ we have:

$$1 \leq DP_{ij}(\overline{y}^k) \qquad \text{and} \qquad DP_{ij}(\overline{y}^k) \underbrace{\frac{x^0}{z^0 - x^0} \frac{(s_z - s_x)}{s_x}}_{\beta} \leq \frac{1}{s(y_i, \overline{y}^k)}. \tag{2.31}$$

As this subcase is such that $s_z > s_x \frac{z^0}{x^0}$, we have $\beta > 1$. Observe that the slope of a given equivalence curve is constant for all $\overline{y} \geq \overline{y}^k$, which implies the second inequality can be bounded above by the slope at $\overline{y}^k$.

If inequalities (2.31) are not met when taking $DP_{ij}(\overline{y}^k) = 1$ for all $y_i, y_j \in [0, z(\overline{y}^k))$ with $y_i < x(\overline{y}^k) < y_j$, then any other value for $DP_{ij}(\overline{y}^k)$ also implies their violation.[58] In other words, if the PGR at $\overline{y}^k$ cannot respect these conditions, no other numerical representation can. On the contrary, if the PGR at $\overline{y}^k$ does respect inequalities (2.31), then the index based on this numerical representation satisfies both *Monotonicity in Income* and *Transfer among Poor*. Indeed, I showed in Step 1 that respecting the condition of Lemma 8 for all $y_i, y_j \in [0, z(\overline{y}^k))$ with $y_i < y_j < x^0$ or with $x^0 \leq y_i < y_j$ at mean income $\overline{y}^k$ was sufficient to respect it for all $\overline{y} \geq \overline{y}^k$.

I show that the PGR at $\overline{y}^k$ respect inequalities (2.31) for all $y_i, y_j \in [0, z(\overline{y}^k))$ with $y_i < x^0 < y_j$ if and only if $s_x \geq \underline{s}_x$. The first of these inequality holds as $DP_{ij}(\overline{y}^k) = 1$. I show that provided $s_x \geq \underline{s}_x$, the second holds as well for the subcase 3.1. The left-hand side of this second inequality does not depend on the specific value taken by y_i and y_j, given they meet $y_i < x^0 < y_j$. The tightest upper bound is obtained when y_i tends to x^0 and hence $s(y_i, \overline{y}^k)$ tends to s_x. Replacing $DP_{ij}(\overline{y}^k)$ by 1 and $s(y_i, \overline{y}^k)$ by s_x yields successively:

$$\frac{x^0}{(z^0 - x^0)} \frac{(s_z - s_x)}{s_x} \leq \frac{1}{x_0},$$

$$\underline{s}_x := s_z - \frac{(z^0 - x^0)}{x^0} \leq s_x.$$

This $\underline{s}_x$ corresponds to the threshold for s_x below which the PGR at the origin leads to an index violating either *Transfer among Poor* or *Monotonicity in Income* (and

[58] The reason is that this definition of DP_{ij} considers its minimal value such that inequality $1 \leq DP_{ij}(\overline{y}^k)$ holds.

hence any other numerical representation as well) and above which the PGR at the origin leads to an index satisfying both properties. This bound is such that $\underline{s}_x < s_x^h$:

$$s_z - \frac{(z^0 - x^0)}{x^0} < \frac{x^0}{z^0} s_z,$$
$$s_z < \frac{z^0}{x^0},$$

which holds as $s_z \leq 1$ and we assumed $x^* < z^0$.

STEP 3: For subcase 3.2, derive the upper bound $\overline{s}_x$ for s_x below which an additive index satisfies the necessary and sufficient conditions of Lemma 8.

Subcase 3.2 is such that $\frac{x^0}{z^0} s_z < s_x < s_z$. Step 1 showed for this case that $DP_{ij}(\overline{y})$ is strictly decreasing in $\overline{y}$. As for Step 2, it is sufficient to check the condition in Lemma 8 at the boundaries:

$$DP_{ij}(\overline{y}^k) \leq \frac{1}{s(y_i, \overline{y}^k)} \qquad \text{and} \qquad 1 \leq DP_{ij}(\overline{y}^k) \underbrace{\frac{x^0}{z^0 - x^0} \frac{(s_z - s_x)}{s_x}}_{\beta}. \tag{2.32}$$

As this subcase is such that $s_z < s_x \frac{z^0}{x^0}$, we have $\beta < 1$. If inequalities (2.32) are not met when taking $DP_{ij}(\overline{y}^k)\beta = 1$ for all $y_i, y_j \in [0, z(\overline{y}))$ with $y_i < x^0 < y_j$ when $\overline{y}$ tends to ∞, then any other value for $DP_{ij}(\overline{y}^k)$ also implies their violation. In other words, if the PGR at $\overline{y}^\infty$ cannot respect these conditions, no other numerical representation of $\succeq_{s_x}$ can.[59] On the contrary, if the PGR at $\overline{y}^\infty$ does respect inequalities (2.31), then the index based on this numerical representation satisfies both *Monotonicity in Income* and *Transfer among Poor*, as explained in Step 2.

I show that the PGR at $\overline{y}^\infty$ respect inequalities (2.32) for all $y_i, y_j \in [0, z(\overline{y}^k))$ with $y_i < x^0 < y_j$ if and only if $s_x \leq \overline{s}_x$. The second of these inequality holds as $DP_{ij}(\overline{y}^k)\beta = 1$. I show that provided $s_x \leq \overline{s}_x$, the first holds as well for subcase 3.2. The tightest upper bound is obtained when y_i tends to x^0 and hence $s(y_i, \overline{y}^k)$ tends to s_x. Replacing $DP_{ij}(\overline{y}^k)$ by $\frac{1}{\beta}$ and $s(y_i, \overline{y}^k)$ by s_x yields successively:

$$\frac{(z^0 - x^0)}{x^0} \frac{s_x}{(s_z - s_x)} \leq \frac{1}{s_x},$$
$$s_x^2 + s_x \frac{x^0}{z^0 - x^0} - s_z \frac{x^0}{z^0 - x^0} \leq 0.$$

This second order equation in s_x has two roots r_- and r_+, one negative and one positive. The images of this parabol are negative between the two roots. The positive root constitutes the threshold $\overline{s}_x := r_+$, given by:

$$\overline{s}_x = \frac{\left(\left(\frac{x^0}{z^0 - x^0}\right)^2 + 4 s_z \frac{x^0}{z^0 - x^0}\right)^{0.5} - \frac{x^0}{z^0 - x^0}}{2}.$$

This $\overline{s}_x$ corresponds to the threshold for s_x above which the PGR at $\overline{y}^\infty$ leads to an index violating either *Transfer among Poor* or *Monotonicity in Income* (and hence

[59] The PGR at $\overline{y}^\infty$ is defined as the numerical representation in the extended FGT family with $\alpha = 1$ granting a degree of priority at $\overline{y}^k$ to absolute over relatively poor agents of $\frac{1}{\beta}$.

any other numerical representation as well) and above which the PGR at $\overline{y}^{\infty}$ leads to an index satisfying both properties. This bound is such that $s_x^h := \frac{x^0}{z^0} s_z < \overline{s}_x$ as successively we have:

$$2\frac{x^0}{z^0}s_z + \frac{x^0}{z^0-x^0} < \left(\left(\frac{x^0}{z^0-x^0}\right)^2 + 4s_z\frac{x^0}{z^0-x^0}\right)^{0.5},$$

$$\left(\frac{x^0}{z^0}s_z\right)^2 < \frac{x^0}{z^0}s_z,$$

which is guaranteed as $s_z \leq 1$ and $x^0 < z^0$. This bound is also such that $\overline{s}_x < s_z$ as successively we have:

$$\left(\left(\frac{x^0}{z^0-x^0}\right)^2 + 4s_z\frac{x^0}{z^0-x^0}\right)^{0.5} < 2s_z + \frac{x^0}{z^0-x^0},$$

$$\frac{x^0}{z^0} < \frac{x^0}{z^0} + s_z,$$

which is guaranteed as $s_z > 0$ as $\bar{s} > 0$. This concludes the proof of Claim 1.

I prove Claim 2, based on arguments exposed in the proof of Claim 1. A direct application of the reasoning proving Claim 1 of Theorem 2 shows that any P having its numerical representation in the extended FGT family satisfies *Monotonicity in Income* only if $\alpha = 1$. Then, as by assumption $s_x \leq \frac{x^0}{z^0}s_z$, the expression of $DP_{ij}(\overline{y})$ for any $y_i < x(\overline{y}) < y_j$ given in (2.30) is strictly increasing in $\overline{y}$ as shown in Step 1 of Claim 1. Therefore, if the numerical representation d of P is the PGR at $\overline{y}^r$ then P satisfies *Transfer among Poor* only if $\overline{y}^r = \overline{y}^k$, that is d is the PGR at the origin. Finally, Step 2 of Claim 1 showed that if $s_x \geq 0$ is such that $s_x \in [\underline{s}_x, s_x^h]$ and if d is the PGR at the origin, then P satisfies both *Monotonicity in Income* and *Transfer among Poor* when $s_x \leq \frac{x^0}{z^0}s_z$. This shows the equivalence of the two statements in Claim 2.

2.9.6 Proof of Theorem 5

The proof is very close to that of Theorem 1. I therefore omit parts that are straightforward modifications in order to emphasize the differences.

Take any median-sensitive endogenous line z and any poverty index P satisfying the five modified axioms. Let $Y^r_{even} := \{y \in Y^r | n(y) \in 2\mathbb{N}\}$ be the subdomain of Y^r containing only income distributions with an even number of dimensions. I prove in Steps 1 to 3 that the additive representation is implied for all $y \in Y^r_{even}$, then in Step 4 I use *Replication Invariance* to extend this result to the whole Y^r.

STEP 1: From a poverty ordering on income distributions to a poverty ordering on distributions of individual poverty.

The definition of the mapping is different. Let the continuous mapping be M :

$Y^r_{even} \to \mathbb{R}^{N'}$, where $N' := \{n \in 2\mathbb{N}+1 | n \geq 1\}$. Let $\succeq^m$ be an EO in $\mathcal{R}_m$ whose unanymous judgments among the poor are respected by P. By modified *Domination among Poor*, such $\succeq^m$ exists. Consider any numerical representation d of $\succeq^m$. For each $(y_i, \overline{y}) \in X$, let $\nu_i := d(y_i, y_m)$. Mapping M is defined for all $y \in Y^r_{even}$ such that

$$M(y) = (\nu_1, \cdots, \nu_{m-1}) := \nu.$$

Observe that if distribution $y \in Y^r_{even}$ has n components, then $M(y)$ has $m-1$ components. The size of distribution ν is taken to be $m-1$ as for all $y \in Y$ and all $i \geq m$ we have $d(y_i, y_m) = 0$ since $y_i \geq z(y_m)$ and are hence omitted.

I show for the continuous mapping defined that $M(Y^r_{even}) = V_d := [0,1]^{N'}$. The domain of images of Y^r_{even} through mapping M is hence a product space: $V_d = \times_{i=1}^{N'}[0,1]_i$. This means that (i) $M(Y^r_{even}) \subseteq V_d$ and (ii) $V_d \subseteq M(Y^r_{even})$, that is for all $\nu \in V_d$ there exists $y \in Y^r_{even}$ such that $M(y) = \nu$. If (i) follows directly from the definition of mapping M, (ii) remains to be proven. Lemma 9 proves that $V_d \subseteq M(Y^r_{even})$.

Lemma 9. *For all $\succeq^m \in \mathcal{R}_m$ and $\nu \in V_d$, there exists $y \in Y^r_{even}$ such that $M(y) = \nu$.*

Proof. Consider any $\nu \in V_d$ and any $g \in \mathbb{R}_{++}$ such that $g \geq z(g)$. For any endogenous line, such a g exists by modified Possibility of Poverty Eradication. I construct $y \in \mathbb{R}_+^{2(n(\nu)+1)}$ such that $y_m = g$ and $M(y) = \nu$. Let y be such that, for all $i \leq n(\nu)$, $y_i := a_i$ defined implicitly by $\nu_i = d(a_i, g)$. If $\nu_i = 0$ take $a_i := g$. By modified Minimal Absolute Concern and the continuity of d, we have that $a_i \in [0, z(g))$ for all $i \leq q$. For all $j \in \{m(y), \ldots, n(y)\}$, take $y_j := g$. This construction implies $y_m = g$. Therefore we have $y \in Y^r_{even}$ since (i) $n(y) = 2(n(\nu)+1)$ is even, (ii) a majority of agents are non-poor as $n(\nu)+2$ agents earn $g \geq z(g)$ and we have $n(\nu)+2 > n(y)/2$. We have by construction $M(y) = \nu$, the desired result. ∎

Again, the poverty ordering $\succeq_{Y^r_{even}}$ on the set of income distributions is associated to an ordering $\succeq_{V_d}$ on distributions of individual poverty, by modified *Domination among Poor*. This ordering $\succeq_{V_d}$ can be represented by a continuous poverty index $P^v : V_d \to \mathbb{R}$.[60] In particular, ordering $\succeq_{V_d}$ is represented by P^ν defined such that for all $\nu \in V_d$ and $y \in Y^r_{even}$ with $M(y) = \nu$, we have $P^\nu(\nu) = P^\nu\big(M(y)\big) = P(y)$.

STEP 2: Index P^ν representing ordering $\succeq_{V_d}$ is additively separable.

We verify that the assumptions of Theorem 1 in Gorman (1968) are all met. This allows deriving the following functional form for the index P^ν, for a given $n(y) \in 2\mathbb{N}$:

$$P^\nu(\nu) = \tilde{F}\left(\sum_{i=1}^{m-1} \tilde{\varphi}(\nu_i)\right) \tag{2.33}$$

where $\tilde{F}$ and $\tilde{\varphi}$ are strictly increasing functions and $m-1 = n(\nu)$.

[60] Mapping M cannot be used to obtain the image of income distributions with *odd* number of dimensions, otherwise the same ν is obtained for two income distributions with different poverty. For example $\nu = (1)$, corresponds to both $y_{even} := (0, g, g, g)$ and $y_{odd} := (0, g, g)$; and the ranking on V_d cannot discriminate these two income distributions having different poverty.

The assumptions required for this Theorem are the following:

Assumption 1: As before.

Assumption 2: As before.

Assumption 3: Let $S := \{[0,1]_1, \cdots, [0,1]_{m-1}\}$ be the set of sectors in V_d and $A \subseteq S$ be any subset of sectors, we have that each A is *separable*. *Separability* means that for all $(u,w), (v,w), (u,t), (v,t) \in V_d$, we have $P^\nu(u,w) \geq P^\nu(v,w) \;\Leftrightarrow\; P^\nu(u,t) \geq P^\nu(v,t)$. Separability is proven in two substeps.

Substep 1: Construct an income distributions associated to each distribution of individual poverty.

Construct $y^1, y^2, y^3, y^4 \in Y^r_{even}$ such that $M(y^1) = (u,w)$, $M(y^2) = (v,w)$, $M(y^3) = (u,t)$, $M(y^4) = (v,t)$, $y^1, y^2, y^3, y^4 \in \mathbb{R}_+^{2(n(u,v)+1)}$ and $y^1_m = y^2_m = y^3_m = y^4_m = g$ with $g \geq z(g)$. Such distributions exist and are constructed following the procedure presented in Lemma 9.

Decompose in subgroups $y^1 = (y^1_A, y^1_B, y^1_C)$, such that subdistributions y^1_A and y^1_B are associated –via the numerical representation d – to the individual poverty subdistributions u and w respectively and y^1_C is the subdistributions containing the income for all $j \in \{m(y), \ldots, n(y)\}$, for whom by construction we have $y_j = g$. By construction, $y^1 \in \mathbb{R}_+^{2(n(u,v)+1)}$ and has hence an *even* number of dimensions. We can hence decompose $y^1_C = (y^1_{C1}, y^1_{C2}) = (g, \ldots, g)$ such that $n(y^1_{C1}) = n(y^1_A) + 1$ and $n(y^1_{C2}) = n(y^1_B) + 1$. Typically, the median in y^1_A is different from the median in y^1_B, which is different from g but our next operations will aim at equalizing median income in subgroups by distributing the agents in subgroup C between the subgroups A and B.

Duplicate y^1 and re-organize the subgroups in a way that equalizes median income in each subgroup with g. Let $x^1 := (y^1, y^1) = (x^1_{A'}, x^1_{B'})$ with $x^1_{A'} := (y^1_A, y^1_A, y^1_{C1}, y^1_{C1})$ and $x^1_{B'} := (y^1_B, y^1_B, y^1_{C2}, y^1_{C2})$, implying both $x^1_{A'}$ and $x^1_{B'}$ have an even number of dimensions. This duplication does not affect the median: $y^1_{m(y^1)} = x^1_{m(x^1)} = g$. Furthermore, we have by construction that both the median of $x^1_{A'}$ and $x^1_{B'}$ are equal to g and $M(x^1_{A'}) = (u,u)$ and $M(x^1_{B'}) = (w,w)$. Observe finally that $M(x^1) = (u,u,w,w,0)$, as $d(g,g) = 0$.

Using the same procedure (decomposition, duplication, reorganization), construct x^2, x^3, x^4 such that:

$$\begin{aligned}
x^1 &= (x^1_{A'}, x^1_{B'}) && \text{with } M(x^1) = (u,u,w,w,0),\\
x^2 &= (x^2_{A'}, x^2_{B'}) && \text{with } M(x^2) = (v,v,w,w,0),\\
x^3 &= (x^3_{A'}, x^3_{B'}) && \text{with } M(x^3) = (u,u,t,t,0),\\
x^4 &= (x^4_{A'}, x^4_{B'}) && \text{with } M(x^4) = (v,v,t,t,0).
\end{aligned}$$

For all $m \in \{1,2,3,4\}$, we have $P(x^m) = P(y^m)$ by *Replication Invariance*. Therefore, proving $P(x^1) \geq P(x^2) \Leftrightarrow P(x^3) \geq P(x^4)$ is equivalent to proving $P^\nu(u,w) \geq P^\nu(v,w) \;\Leftrightarrow\; P^\nu(u,t) \geq P^\nu(v,t)$. For notational simplicity, drop the symbols $'$ to name the new subgroups A' and B' as the old ones.

Substep 2: Prove $P(x_A^1, x_B^1) \geq P(x_A^2, x_B^2) \Leftrightarrow P(x_A^3, x_B^3) \geq P(x_A^4, x_B^4)$.

Our income distributions are constructed such that $P(x_A^1) = P(x_A^3)$, $P(x_A^2) = P(x_A^4)$, $P(x_B^1) = P(x_B^2)$ and $P(x_B^3) = P(x_B^4)$ by *Domination among Poor*. By assumption we have $P(x^1) \geq P(x^2)$. As $P(x_B^1) = P(x_B^2)$, we have that $P(x_A^1) \geq P(x_A^2)$ by *Weak Subgroup Consistency* (remember all our subgroups have their median equal to g).

Then, $P(x_A^1) \geq P(x_A^2)$ together with $P(x_A^1) = P(x_A^3)$ and $P(x_A^2) = P(x_A^4)$ imply $P(x_A^3) \geq P(x_A^4)$. Since $P(x_B^3) = P(x_B^4)$, this implies $P(x^3) \geq P(x^4)$. Two cases can arise.

- Case 1: $P(x_A^3) > P(x_A^4)$.
  As $P(x_B^3) = P(x_B^4)$, we obtain $P(x_A^3, x_B^3) > P(x_A^4, x_B^4)$, by *Weak Subgroup Consistency*. This case is hence such that $P(x^3) \geq P(x^4)$ as desired.

- Case 2: $P(x_A^3) = P(x_A^4)$.
  I show by contradiction that $P(x^3) \geq P(x^4)$. Assume we have $P(x_A^3, x_B^3) < P(x_A^4, x_B^4)$. As $P(x_A^3) = P(x_A^4)$, *Weak Subgroup Consistency* implies that $P(x_A^3, x_B^3, x_A^4) < P(x_A^4, x_B^4, x_A^3)$. Again, as $P(x_B^3) = P(x_B^4)$, we obtain $P(x_A^3, x_B^3, x_A^4, x_B^4) < P(x_A^4, x_B^4, x_A^3, x_B^3)$. This is a contradiction as the two distributions have identical poverty by *Symmetry*.

The two cases lead to $P(x^3) \geq P(x^4)$, which proves separability. All assumptions of Theorem 1 in Gorman (1968) are met.

STEP 3: Show functions $\tilde{F}$ and $\tilde{\varphi}$ do not depend on the number n of agents.

Theorem 1 in Gorman (1968) is valid for a fixed number of potentially poor agents $n(\nu)$. I modify the proof of Foster and Shorrocks (1991) in order to prove these functions are independent of n. When $n(\nu)$ is allowed to vary – $n(\nu)$ will be denoted n below – (2.33) must be written:

$$P^{\nu}(\nu) = \tilde{F}_n\left(\sum_{i=1}^{n} \tilde{\varphi}_n(\nu_i)\right)$$

Step 3.1: Define transformations of $\tilde{F}_n$ and $\tilde{\varphi}_n$ for normalization purposes.
Let F_n and φ_n be the following transformations of $\tilde{F}_n$ and $\tilde{\varphi}_n$:

$$\varphi_n(\nu_i) = 2(n+1)[\tilde{\varphi}_n(\nu_i) - \tilde{\varphi}_n(0)],$$
$$F_n(x) = \tilde{F}_n[x + n\tilde{\varphi}_n(0)].$$

These transformations imply successively:

$$F_n\left(\frac{1}{2(n+1)}\sum_{i=1}^{n}\varphi_n(\nu_i)\right) = F_n\left(\frac{2(n+1)}{2(n+1)}\sum_{i=1}^{n}[\tilde{\varphi}_n(\nu_i) - \tilde{\varphi}_n(0)]\right),$$
$$= \tilde{F}_n\left(\sum_{i=1}^{n}[\tilde{\varphi}_n(\nu_i) - \tilde{\varphi}_n(0)] + n\tilde{\varphi}_n(0)\right).$$

This yields

$$P^{\nu}(\nu) = F_n\left(\frac{1}{2(n+1)}\sum_{i=1}^{n}\varphi_n(\nu_i)\right),$$

where $\varphi_n(0) = 0$ and by the definition of mapping M, we have $2(n+1) = n(y)$. As any agent $j \in \{m(y), \dots, n(y)\}$ is non-poor in Y^r, we have $d(y_j, y_m) = 0$. By slightly abusing notation (by introducing the zero individual poverty of those non-poor agents at the end of the distribution ν), we obtain for all even $n(y)$:

$$P^{\nu}(\nu) = F_n\left(\frac{1}{n(y)}\left(\sum_{i=1}^{m(y)-1}\varphi_n(\nu_i) + \sum_{i=m(y)}^{n(y)}\varphi_n(0)\right)\right) \tag{2.34}$$

$$= F_n\left(\frac{1}{n(y)}\sum_{i=1}^{n(y)}\varphi_n(\nu_i)\right) \tag{2.35}$$

with F_n and φ_n continuous, strictly increasing and $\varphi_n(0) = 0$.

Step 3.2: Use *Replication Invariance* to prove functions F_n and φ_n do not depend on n.

Same as before.

STEP 4: The additively separable expression obtained for all $y \in Y^r_{even}$ is valid for all $y \in Y^r_{odd}$.

Consider any $y \in Y^r_{odd}$ and its duplication $x := (y, y) \in Y^r_{even}$. By *Replication Invariance*, we have $P(y) = P(x)$, which means that the mathematical expression of any $y \in Y^r_{odd}$ also take the additively separable form of (2.35) as:

$$P(x) = F\left(\frac{1}{n(x)}\sum_{i=1}^{n(x)} d\big(x_i, x_m\big)\right) = F\left(\frac{1}{2n(y)}\sum_{i=1}^{n(y)} 2d\big(y_i, y_m\big)\right) = P(y),$$

where $d\big(y_i, y_m\big) = d\big(x_{2i}, x_m\big)$ for all $i \leq n(y)$ as $y_m = x_m$ and $y_i = x_{2i}$. This completes the proof as $Y^r = Y^r_{odd} \cup Y^r_{even}$.

2.9.7 Proof of Theorem 6

Let z be any median-sensitive line with $z^0 > 0$. Let P be any additive poverty index based on an absolute-homothetic EO below z.

I don't provide a complete proof showing that the second statement implies the first. The intuition is the following. For any additive index based on an absolute-homothetic EO below a line z meeting the second statement, *Monotonicity in Income* is implied by *Domination among Poor*. As additive indices respect *Domination among Poor*, such P satisfies *Monotonicity in Income*.

I prove by contraposition that the first statement implies the second. As by assumption $z^0 > 0$, we have $y_m^* > 0$. Assume there exists $y_m^1 < y_m^*$ with $s\left(y_m^1\right) > 0$. I construct an $y \in Y^p$ at which a violation of *Monotonicity in Income* arise for a particular increment. Let $y_m := y_m^1$ and let y_{m+1} be constructed such that $y_{m+1} \leq z(y_m^1)$ and $y_{m+1} - y_m = \epsilon > 0$. The numerical representation of P is d. Let $\Delta := d(y_m, y_m^1) - d(y_m + \epsilon, y_m^1 + \epsilon)$ be the individual poverty gain obtained by the median agent when her income is increased by the increment ϵ. Observe that the increase in median income with ϵ does not depend on the number of agents, contrary to the increase in mean income. By modified Strict Monotonicity in Income and Translation Monotonicity, we have that $\Delta \geq 0$.

Consider income level $a < z(y_m^1)$ with $s(a, y_m^1) > 0$. Such an income a exists as the EO is absolute-homothetic and $s\left(y_m^1\right) > 0$. Let $\delta := d(a, y_m^1 + \epsilon) - d(a, y_m^1)$ be the individual poverty loss obtained by an agent earning a when the income of the median agent is increased by ϵ. We have $\delta > 0$ since $s(a, y_m^1) > 0$ and numerical representation d is strictly decreasing in equivalence levels by modified Strict Monotonicity in Income.

Let n^a be the number of agents earning income a in the income distribution y. If $n^a > \frac{\Delta}{\delta}$, giving an additional ϵ to the median agent in distribution y strictly increases poverty. Let y' be obtained from y when median agent earns an extra ϵ. As P is an additive index, we have:

$$
\begin{aligned}
P(y) - P(y') =& \frac{1}{n} \sum_{i=1}^{n} \left(d\big(y_i, y_m\big) - d\big(y_i', y_m'\big) \right), \\
=& \frac{1}{n} \left(\Delta - \left(\sum_{j \neq m} d\big(y_j, y_m^1 + \epsilon\big) - d\big(y_j, y_m^1\big) \right) \right), \\
=& \frac{1}{n} \left(\Delta - n^a \delta - A \right),
\end{aligned}
$$

where term A stands for the sum of individual poverty losses obtained by agents different than the median agent *and* not earning a. We have $A \geq 0$ as the EO is absolute-homothetic and $\epsilon > 0$, implying that the individual poverty of all poor agents except the median agent cannot decrease when passing from y to y'. As a result, if $n^a > \frac{\Delta}{\delta}$ then we have $P(y) - P(y') < 0$, which violates *Monotonicity in Income*. There exists such $y \in Y^p$ since the number of agents in distributions belonging to Y^p is not bounded above.

2.9.8 Partial means

I extend here Theorem 1 and Theorem 2 for the class of lower partial means. The case $x = 100$ corresponds to the mean.

Monotonicity in Income has different implications when imposed on mean-sensitive or median-sensitive poverty lines. For mean-sensitive lines, Theorems 2 and 3 show this axiom forces the index to be close to the PGR at the origin. Unlike the median, the mean as well as lower partial means are influenced by the income of all poor agents.[61]

The definition of a lower partial mean given in (2.36) is in two parts since I consider finite income distributions and there is hence no guarantee that $\frac{x}{100}n$ be a natural number. For the sake of notational simplicity, let x denote a *fraction* rather than a percentage, that is $x \in (0, 1]$.

Definition 17 (Lower partial mean).
The income standard $f^{\ell pm} : \mathbb{R}^n_+ \to \mathbb{R}_+$ *is a lower partial mean if*

$$f^{\ell pm}(y) := \begin{cases} \frac{1}{xn} \sum_{i=1}^{xn} y_i & \text{if } xn \in \mathbb{N}, \\ \\ \frac{1}{xn} \left(\sum_{i=1}^{r} y_i + (xn - r) y_{r+1} \right) & \text{otherwise,} \end{cases} \tag{2.36}$$

where $x \in \{a \in \mathbb{Q} | 0 < a \leq 1\}$ *and* $r := \max_{a \in \mathbb{N}} a \leq xn$.

Again, changing the income standard requires modifying several definitions. I present here only the main non-straightforward modifications. The domain of income distributions ranked by poverty indices with partial-mean-sensitive lines depend on the particular partial mean used. This restriction is necessary for the characterization of aditive indices. Let $f^{\ell pm}_x$ be the lower partial mean ignoring the $100(1-x)\%$ richest agents. Any income distribution in the set Y^f contains at least $100(1-x)\%$ non-poor agents:

$$Y^f := \{y \in \mathbb{R}^n_+ | z(f^{\ell pm}_x(y)) \leq y_r\}.$$

Poverty indices are based on an equivalence ordering $\succeq^f$ on the set of poor bundles X_p defined as:

$$X_p := \{(y_i, f^{\ell pm}_x(y)) \in X | y_i < z(f^{\ell pm}_x(y))\},$$

where $X := \mathbb{R}_+ \times \mathbb{R}_{++}$. Restrictions to endogenous lines and EOs are easily modified except maybe modified Translation monotonicity:

EO restriction 6 (modified Translation monotonicity).
For all $(y_i, f^{\ell pm}_x(y)) \in X_p$ *and* $a > 0$, *we have* $\big(y_i + a, f^{\ell pm}_x(y + a\mathbf{1_n})\big) \succeq \big(y_i, f^{\ell pm}_x(y)\big)$.[62]

All axioms on poverty indices are easily modified except maybe *modified Weak Subgroup Consistency*:

Social ordering axiom 21 (*modified Weak Subgroup Consistency*).
For all $y^1, y^2, y^3, y^4 \in Y^f$ *such that* $n(y^1) = n(y^3)$, $n(y^2) = n(y^4)$, $f^{\ell pm}_x(y^1) = f^{\ell pm}_x(y^2) = f^{\ell pm}_x(y^1, y^2)$ *and* $f^{\ell pm}_x(y^3) = f^{\ell pm}_x(y^4) = f^{\ell pm}_x(y^3, y^4)$, *if* $P(y^1) > P(y^3)$ *and* $P(y^2) = P(y^4)$, *then* $P(y^1, y^2) > P(y^3, y^4)$.

Such modifications allow characterizing additive poverty indices with partial-mean-sensitive lines for the domain of income distributions containing at least $100(1-x)\%$ of non-poor agents. For notational simplicity, the lower partial mean $f^{\ell pm}_x$ is denoted f.

[61] The incomes of all poor agents influence the lower partial mean if the percentage of poor agents is lower than x.

[62] Where $\mathbf{1_n}$ denotes a n-dimensional distribution of ones. Giving an equal increment to all agents cannot increase the individual poverty of a poor agent.

Theorem 9 (Characterization of additive partial-mean-sensitive indices)**.**
Let f be a lower partial mean. Let $P : Y^f \to \mathbb{R}$ be a poverty index based on a f-sensitive poverty line. The following two statements are equivalent.

1. *P is ordinally equivalent to an index $P' : Y^f \to [0,1]$ with*
$$P'(y) = \frac{1}{n}\sum_{i=1}^{n} d(y_i, f(y)), \tag{2.37}$$
where d is a numerical representation of an EO in $\mathcal{R}$.

2. *P satisfies the modified versions of Domination among Poor, Weak Subgroup Consistency, Symmetry, Continuity and Replication Invariance.*

Proof. The proof is very close to the proof of Theorem 1. I therefore omit parts that are straightforward modifications in order to emphasize differences. Again, I just prove that statement two implies statement one.

Take any partial-mean-sensitive endogenous line z and any poverty index P satisfying the five modified axioms. By assumption we have $x \in \mathbb{Q}$ so x can be expressed as $x = \frac{a}{b}$ with $a, b \in \mathbb{N}$ and their greatest common divider $GCD(a,b) = 1$. Let $Y^f_{b\mathbb{N}}$ be the subset of income distributions in Y^f for which xn belongs to the natural:
$$Y^f_{b\mathbb{N}} := \left\{ y \in Y^f | n(y) \in b\mathbb{N} \right\}.$$
I prove in Steps 1 to 3 that the additive representation is implied for all $y \in Y^f_{b\mathbb{N}}$, then in Step 4 I use *Replication Invariance* to extend it to the wole Y^f.

STEP 1: From a poverty ordering on income distributions to a poverty ordering on distributions of individual poverty.

The definition of the mapping is different. Let the continuous mapping be $M : Y^f_{b\mathbb{N}} \to \mathbb{R}^{N'}$, where $N' := \{n \in \mathbb{N} \mid n+1 \in a\mathbb{N} \text{ and } n \geq 1\}$. Let $\succeq^f$ be an EO in $\mathcal{R}_f$ whose unanymous judgments among the poor are respected by P. By modified *Domination among Poor*, such $\succeq^f$ exists. Consider any numerical representation d of $\succeq^f$. For each $(y_i, f(y)) \in X$, let $\nu_i := d(y_i, f(y))$. Mapping M is defined for all $y \in Y^f_{b\mathbb{N}}$ such that
$$M(y) = (\nu_1, \cdots, \nu_{xn-1}) := \nu.$$
Observe that if distribution $y \in Y^f_{b\mathbb{N}}$ has n components, then $M(y)$ has $xn-1$ components. The size of distribution ν is taken to be $xn-1$ as for all $y \in Y$ and all $i \geq xn$ we have $d(y_i, f(y)) = 0$ since $y_i \geq z(f(y))$ and are hence omitted. I show for the continuous mapping defined that $M(Y^f_{b\mathbb{N}}) = V_d := [0,1]^{N'}$. The domain of images of $Y^f_{b\mathbb{N}}$ through mapping M is hence a product space: $V_d = \times_{i=1}^{N'}[0,1]_i$. This means that (i) $M(Y^f_{b\mathbb{N}}) \subseteq V_d$ and (ii) $V_d \subseteq M(Y^f_{b\mathbb{N}})$, that is for all $\nu \in V_d$ there exists $y \in Y^f_{b\mathbb{N}}$ such that $M(y) = \nu$. If (i) follows directly from the definition of mapping M, (ii) remains to be proven. Lemma 10 proves that $V_d \subseteq M(Y^f_{b\mathbb{N}})$.

Lemma 10. *For all $\succeq^f \in \mathcal{R}_f$ and $\nu \in V_d$, there exists $y \in Y^f_{b\mathbb{N}}$ with $M(y) = \nu$.*

Proof. Consider any $\nu \in V_d$ and any $g \in \mathbb{R}_{++}$ such that $g \geq z(g)$. For any endogenous line, such a g exists by modified Possibility of Poverty Eradication. I construct $y \in Y^f_{b\mathbb{N}}$ such that $f(y) = g$ and $M(y) = \nu$. By the definition of V_d, we have that $\frac{n(\nu)+1}{x} \in b\mathbb{N}$. Let y be such that, for all $i \leq xn - 1$, $y_i := c_i$ defined implicitly by $\nu_i = d(c_i, g)$. If $\nu_i = 0$ take $c_i := g$. By modified Minimal Absolute Concern and the continuity of d, we have that $c_i \in [0, z(g))$ for all $i \leq q$. Take then y_{xn} such that $f(y) = \frac{1}{xn}\sum_{i=1}^{xn} y_i = g$. I prove now that such a y_{xn} exists and is such that $y_{xn} \geq z(g)$. Consider y' with $n(y') = n$ and for which $y'_i := y_i$ for all $i \leq xn - 1$ and $y'_{xn} := g$. By construction, we have $y'_i \leq g$ for all $i \in \{1, \ldots, xn(y')\}$ as $g \geq z(g)$. Therefore we have $f(y') \leq g$. Now, there exists $c \geq g$ such that if $y_n := c$ then $f(y) = g$. Since $c \geq g$ and $g \geq z(g)$ by assumption, we have that agent xn is non-poor. Take finally $y_i := y_{xn}$ for all $i \in \{xn + 1, \ldots, n\}$. This ensures these agents are non-poor and have a weakly higher income than agent xn and hence $y \in Y^f_{b\mathbb{N}}$. By construction of y, we have $f(y) = g$ and $M(y) = \nu$. ■

Again, the poverty ordering $\succeq_{Y^f_{b\mathbb{N}}}$ on the set of income distributions is associated to an ordering $\succeq_{V_d}$ on distributions of individual poverty, by modified *Domination among Poor*. This ordering $\succeq_{V_d}$ can be represented by a continuous poverty index $P^v : V_d \to \mathbb{R}$. In particular, ordering $\succeq_{V_d}$ is represented by P^ν defined such that for all $\nu \in V_d$ and $y \in Y^f_{b\mathbb{N}}$ with $M(y) = \nu$, we have $P^\nu(\nu) = P^\nu\big(M(y)\big) = P(y)$.

STEP 2: Index P^ν representing ordering $\succeq_{V_d}$ is additively separable.

We verify that the assumptions of Theorem 1 in Gorman (1968) are all met. This allows deriving the following functional form for the index P^ν, for a given $n(y) \in b\mathbb{N}$:

$$P^\nu(\nu) = \tilde{F}\left(\sum_{i=1}^{xn-1} \tilde{\varphi}(\nu_i)\right) \tag{2.38}$$

where $\tilde{F}$ and $\tilde{\varphi}$ are strictly increasing functions.

The assumptions required for this Theorem are the following:

Assumption 1: As before.

Assumption 2: As before.

Assumption 3: Let $S := \{[0,1]_1, \cdots, [0,1]_{n(\nu)}\}$ be the set of sectors in V_d and $A \subseteq S$ be any subset of sectors, we have that each A is *separable*. *Separability* means that for all $(u, w), (v, w), (u, t), (v, t) \in V_d$, we have $P^\nu(u, w) \geq P^\nu(v, w) \Leftrightarrow P^\nu(u, t) \geq P^\nu(v, t)$. Separability is proven in two substeps.

Substep 1: Construct the income distributions associated to these distributions of individual poverty.

Construct $y^1, y^2, y^3, y^4 \in \mathbb{R}_+^{\frac{n(u,w)+1}{x}}$ such that $M(y^1) = (u, w)$, $M(y^2) = (v, w)$, $M(y^3) = (u, t)$, $M(y^4) = (v, t)$, $y^1, y^2, y^3, y^4 \in Y^f_{b\mathbb{N}}$ and $f(y^1) = f(y^2) = f(y^3) =$

$f(y^4) = g$ with $g \geq z(g)$. Such distributions exist in $Y^f_{b\mathbb{N}}$ and are constructed following the procedure presented in Lemma 10.

Decompose in subgroups $y^1 = (y^1_A, y^1_B, y^1_C)$, such that subdistributions y^1_A and y^1_B are associated to the individual poverty subdistributions u and w respectively and y^1_C is the subdistributions containing the income for all $j \in \{xn, \ldots, n\}$, for whom by construction we have $y_j \geq z(g)$.

By construction we have $n(y^1) = \frac{n(u,w)+1}{x}$ and as $\frac{n(\nu)+1}{x} \in b\mathbb{N}$, $xn(y^1) \in \mathbb{N}$. Typically, we have $f(y^1_A) \neq f(y^1_B) \neq g$, but next operations equalize the reference statistic in new subgroups A' and B' of a k-replication of y^1 by distributing the agents in the k-replication of subgroup C between the k-replications of subgroups A and B and modifying income among some non-poor agents.

Let $s^1 := (y^1, \ldots, y^1)$ be the k-replication of y^1 with $k := b(n(u) + n(w))$. Reorganize the k-subgroups A, B and C in a way to obtain two subgroups A' and B': $s^1 = (s^1_{A'}, s^1_{B'})$ such that $\frac{n(s^1_{A'})}{n(s^1)} = \frac{n(u)}{n(u)+n(w)}$ and all agents associated to the k-replication of subgroup A are in subgroup A'. Given $k = b(n(u) + n(w))$ and $x = \frac{a}{b}$ with $a, b \in \mathbb{N}$, this equality can be obtained with

- $n(s^1_{A'}) \in \mathbb{N}$,
- $xn(s^1_{A'}) \in \mathbb{N}$.

since $bx = a \in \mathbb{N}$ and $xn(y^1) \in \mathbb{N}$. The numbers $n(s^1_{A'})$ and $xn(s^1_{A'})$ belong to the naturals as we have

$$n(s^1_{A'}) = k\left(n(u) + \frac{n(u)}{n(u)+n(w)}[n(y^1) - n(u) - n(w)]\right) = bn(u)n(y^1),$$
$$n(s^1_{B'}) = k\left(n(w) + \frac{n(w)}{n(u)+n(w)}[n(y^1) - n(u) - n(w)]\right) = bn(w)n(y^1),$$

which are such that $n(s^1_{A'}) + n(s^1_{B'}) = n(s^1) = kn(y^1)$. Income distribution $s^1_{A'}$ contains at least $bn(u)$ non-poor agents *and* whose income is taken into account by the lower partial mean f when computing $f(s^1_{A'})$.[63] Accordingly, there are at least $bn(w)$ non-poor agents *and* whose income is taken into account by f when computing $f(s^1_{B'})$.

Construct $s^{1*} = (s^{1*}_{A'}, s^{1*}_{B'})$ from s^1 in such a way that $f(s^{1*}_{A'}) = f(s^{1*}_{B'}) = f(s^{1*}) = g$ and $\big(s^{1*}_i, g)\big) \sim \big(s^1_i, g\big)$ for all $i \in \{1, \ldots, k(n(u) + n(v))\}$. In order to construct s^{1*}, take

- $s^{1*}_i := s^1_i$ for all $i \leq k(n(u) + n(v))$,
- for all agents $\ell \in \{kn(u) + 1, \ldots, kn(u) + bn(u)\}$ in $s^{1*}_{A'}$:

$$s^{1*}_\ell := \frac{xn(s^{1*}_{A'})g - \sum_{i=1}^{kn(u)} s^1_{iA'}}{bn(u)},$$

[63] The number $bn(u)$ is obtained from $xn(s^1_{A'}) - kn(u)$ which is equal to $bn(u)xn(y^1) - b(n(u) + n(w))n(u)$.

- for all $j \in \{kn(w)+1,\dots,kn(w)+bn(w)\}$ in $s^{1*}_{B'}$
$$s^{1*}_j := \frac{xn(s^{1*}_{B'})g - \sum_{i=1}^{kn(w)} s^1_{iB'}}{bn(w)},$$
- $s^{1*}_i := max(s^{1*}_\ell, s^{1*}_j)$ for all $i \in \{xn(s^{1*})+1,\dots,n(s^{1*})\}$,

where both $s^{1*}_\ell \geq g$ and $s^{1*}_j \geq g$, which implies those agents are non-poor. We have by construction that $f(s^{1*}_{A'}) = g$ and $f(s^{1*}_{B'}) = g$.

Observe now that we still have $f(s^{1*}) = g$ as

$$f(s^{1*}) = \frac{1}{xn(s^{1*})}\sum_{i=1}^{xn(s^{1*})} s^{1*}_i = \frac{1}{xn(s^{1*}_{A'}) + xn(s^{1*}_{B'})}\left(\sum_{i=1}^{xn(s^{1*}_{A'})} s^{1*}_{iA'} + \sum_{i=1}^{xn(s^{1*}_{B'})} s^{1*}_{iB'}\right)$$
$$= \frac{xn(s^{1*}_{A'})g + xn(s^{1*}_{B'})g}{xn(s^{1*}_{A'}) + xn(s^{1*}_{B'})} = g.$$

By construction we have

$$M(s^{1*}_{A'}) = (u,\dots,u,0,\dots,0)\,,$$
$$M(s^{1*}_{B'}) = (w,\dots,w,0,\dots,0)\,,$$

where $M(s^{1*}_{A'})$ contains k subdistributions u and $bn(u)-1$ zeros; while $M(s^{1*}_{B'})$ contains k subdistributions w and $bn(w)-1$ zeros.

Using the same procedure (decomposition, k-replication, reorganization), construct s^{2*}, s^{3*}, s^{4*} such that:

$$s^{1*} = (s^{1*}_{A'}, s^{1*}_{B'}) \qquad \text{with } M(s^{1*}) = \Big(u,\dots,u,w,\dots,w,0,\dots,0\Big),$$
$$s^{2*} = (s^{2*}_{A'}, s^{2*}_{B'}) \qquad \text{with } M(s^{2*}) = \Big(v,\dots,v,w,\dots,w,0,\dots,0\Big),$$
$$s^{3*} = (s^{3*}_{A'}, s^{3*}_{B'}) \qquad \text{with } M(s^{3*}) = \Big(u,\dots,u,t,\dots,t,0,\dots,0\Big),$$
$$s^{4*} = (s^{4*}_{A'}, s^{4*}_{B'}) \qquad \text{with } M(s^{4*}) = \Big(v,\dots,v,t,\dots,t,0,\dots,0\Big).$$

where the number of zeros in $M(s^{1*})$ is equal to $b(n(u)+n(w))-1$. For all $m \in \{1,2,3,4\}$, we have $P(s^m) = P(y^m)$ by *Replication Invariance*. As by construction $(s^{m*}_i, g) \sim (s^m_i, g)$ for all $i \leq q(s^{m*}) = q(s^m)$, we have $P(s^{m*}) = P(s^m)$ by *Domination among Poor*. Therefore, proving $P(s^{1*}) \geq P(s^{2*}) \Leftrightarrow P(s^{3*}) \geq P(s^{4*})$ is equivalent to proving $P(y^1) \geq P(y^2) \Leftrightarrow P(y^3) \geq P(y^4)$, which is equivalent to proving $P^\nu(u,w) \geq P^\nu(v,w) \Leftrightarrow P^\nu(u,t) \geq P^\nu(v,t)$. For notational simplicity, drop the symbols * and $'$ to name the new distributions s^{m*} and subgroups A' and B' as the old ones.

Substep 2: Prove $P(s^1_A, s^1_B) \geq P(s^2_A, s^2_B) \Leftrightarrow P(s^3_A, s^3_B) \geq P(s^4_A, s^4_B)$.

Income distributions are constructed such that $s^1_A, s^1_B, s^2_A, s^2_B, s^3_A, s^3_B, s^4_A, s^4_B \in Y^f_{b\mathbb{N}}$, $P(s^1_A) = P(s^3_A)$, $P(s^2_A) = P(s^4_A)$, $P(s^1_B) = P(s^2_B)$ and $P(s^3_B) = P(s^4_B)$ by *Domination among Poor*. The proof is the same as before.

All assumptions of Theorem 1 in Gorman (1968) are met.

STEP 3: Show functions $\tilde{F}$ and $\tilde{\varphi}$ do not depend on the number n of agents.

Theorem 1 in Gorman (1968) is valid for a fixed number of potentially poor agents $n(\nu)$. I modify the proof of Foster and Shorrocks (1991) in order to prove these functions are independent of n. When n is allowed to vary – but still respecting $\frac{n+1}{x} \in b\mathbb{N}$ – (2.38) must be written:

$$P^{\nu}(\nu) = \tilde{F}_n\left(\sum_{i=1}^{n} \tilde{\varphi}_n(\nu_i)\right)$$

Step 3.1: Define transformations of $\tilde{F}_n$ and $\tilde{\varphi}_n$ for normalization purposes.
Let F_n and φ_n be the following transformations of $\tilde{F}_n$ and $\tilde{\varphi}_n$:

$$\varphi_n(\nu_i) = \frac{n+1}{x}[\tilde{\varphi}_n(\nu_i) - \tilde{\varphi}_n(0)],$$
$$F_n(X) = \tilde{F}_n[X + n\tilde{\varphi}_n(0)].$$

These transformations imply successively:

$$\begin{aligned} F_n\left(\frac{x}{n+1}\sum_{i=1}^{n}\varphi_n(\nu_i)\right) =& F_n\left(\frac{x}{n+1}\frac{n+1}{x}\sum_{i=1}^{n}[\tilde{\varphi}_n(\nu_i) - \tilde{\varphi}_n(0)]\right), \\ =& \tilde{F}_n\left(\sum_{i=1}^{n}[\tilde{\varphi}_n(\nu_i) - \tilde{\varphi}_n(0)] + n\tilde{\varphi}_n(0)\right). \end{aligned}$$

This yields

$$P^{\nu}(\nu) = F_n\left(\frac{x}{n+1}\sum_{i=1}^{n}\varphi_n(\nu_i)\right),$$

where $\varphi_n(0) = 0$ and by the definition of N', we have $\frac{n+1}{x} \in \mathbb{N}$.

As any $j \in \{xn(y), \ldots, n(y)\}$ is non-poor in Y^f, we have $d(y_j, g) = 0$. Therefore, we obtain that for all $n(y)$ for which $xn(y) \in \mathbb{N}$, by slightly abusing notation (by introducing the zero individual poverty of those non-poor agents at the end of the distribution ν):

$$\begin{aligned} P^{\nu}(\nu) &= F_n\left(\frac{1}{n(y)}\left(\sum_{i=1}^{xn(y)-1}\varphi_n(\nu_i) + \sum_{i=xn(y)}^{n(y)}\varphi_n(0)\right)\right) \\ &= F_n\left(\frac{1}{n(y)}\sum_{i=1}^{n(y)}\varphi_n(\nu_i)\right) \end{aligned}$$

with F_n and φ_n continuous, strictly increasing and $\varphi_n(0) = 0$.

Step 3.2: Use *Replication Invariance* to prove functions F_n and φ_n do not depend on n.

Same as before.

STEP 4: The additively separable expression obtained for all $y \in Y^f_{b\mathbb{N}}$ is valid for all $y \in Y^f \backslash Y^f_{b\mathbb{N}}$.

Consider any $y \in Y^f$ and $s := (y, \dots, y)$ a k-replication of y with $s \in Y^f_{b\mathbb{N}}$. We have $f(y) = f(s)$ as

$$
\begin{aligned}
f(s) =& \frac{1}{xn(s)} \sum_{i=1}^{xn(s)} s_i, \\
=& \frac{1}{kxn} \left(k \sum_{i=1}^{r} y_i + k\big(xn - r\big) y_{r+1} \right), \\
=& \frac{1}{xn} \left(\sum_{i=1}^{r} y_i + \big(xn - r\big) y_{r+1} \right) = f(y),
\end{aligned}
$$

where $r := \max_{c \in \mathbb{N}} c \leq xn$ and $k\big(xn - r\big) \in \mathbb{N}$.

The mathematical expression of $P(s)$ takes the additive separable form. By *Replication Invariance*, we have $P(y) = P(s)$, which means that the mathematical expression of any $y \in Y^f$ also take the same additive separable form as:

$$
P(s) = \frac{1}{n(s)} \sum_{i=1}^{n(s)} d\big(s_i, f(s)\big) = \frac{1}{kn} \sum_{i=1}^{n} k\ d\big(y_i, f(y)\big) = P(y),
$$

where $d\big(y_i, f(y)\big) = d\big(s_{ki}, f(s)\big)$ for all $i \leq n$ as $f(y) = f(s)$ and $y_i = s_{ki}$.

■

More interestingly, next result shows that the modified version of *Monotonicity in Income* has equivalent implications to those derived when the mean is used as income standard. Observe that on the domain of income distribution Y^f, balanced transfers among poor agents never affect lower partial means.

Theorem 10 (Poverty Gap Ratio for partial-mean-sensitive lines). *Let f be a lower partial mean. Let z be a monotonic f-sensitive poverty line. Let P be an additive index based on an absolute-homothetic EO below z with a numerical representation in the extended FGT family.*

1. *P satisfies modified Monotonicity in Income only if $\alpha = 1$.*
2. *P satisfies modified Monotonicity in Income and modified Transfer among Poor if and only if $\alpha = 1$ and $f^r = 0$,*
   that is d is the PGR at the origin.

Proof. The proof of both claims relies on a modification of Lemma 6 for partial means. This modification provides a necessary condition for Claim 1 and a sufficient condition for Claim 2 under which modified *Monotonicity in Income* is satisfied by additive indices. The modified definitions for the notions of degree of priority and slope are straightforward:

Definition 18 (Degree of Priority partial means). $DP_{ij}(f(y)) := \frac{\partial_1 d(y_i, f(y))}{\partial_1 d(y_j, f(y))}$

Definition 19 (Slope at $(y_i, f(y))$). $s(y_i, f(y)) := -\frac{\partial_2 d(y_i, f(y))}{\partial_1 d(y_i, f(y))}$

Lemma 11. *An additive poverty index based on an absolute-homothetic EO satisfies modified Monotonicity in Income:*

1. *(sufficient condition) if for all $y \in Y^f$ and $y_i, y_j < z\big(f(y)\big)$, we have:*
$$s\big(y_j, f(y)\big) \leq DP_{ij}(f(y)) \tag{2.39}$$
2. *(necessary condition) only if for all $y \in Y^f$ such that there exists $g > 0$ with $f(y) = g$ and $z(g) \leq g$, and all $y_i, y_j < z\big(f(y)\big)$, inequality (2.39) holds.*

Proof. Let f be a lower partial mean. Consider any additive index P based on any f-sensitive line, EO in $\mathcal{R}_f$ and numerical representation d. Monotonicity in Income requires that for all $y \in Y$ and $i \leq q$ we have $\partial_i P(y) \leq 0$. By the additively separable form of P, we obtain by chain derivation:
$$\partial_1 d(y_i, f(y)) + \sum_{j=1}^{n} \partial_2 d(y_j, f(y))\ \partial_i f(y) \leq 0 \tag{2.40}$$
From the definition of lower partial means, we have $\partial_i f(y) = \frac{1}{xn}$.[64] From the definition of $s(y_j, f(y))$, we get $\partial_2 d(y_j, f(y)) = -\partial_1 d(y_j, f(y))\ s(y_j, f(y))$ for all $(y_j, f(y)) \in X$. Inequality (2.40) becomes:
$$\underbrace{\partial_1 d(y_i, f(y)) - \frac{1}{xn} \sum_{j=1}^{n}\ \partial_1 d(y_j, f(y))\ s(y_j, f(y)) \leq 0}_{L_{2.41}} \tag{2.41}$$
In the remainder of the proof, inequality (2.41) is shown to imply the necessary and sufficient conditions linked to (2.39). Inequality (2.39) can be rewritten:
$$\underbrace{\partial_1 d(y_i, f(y)) - \partial_1 d(y_j, f(y))\ s(y_j, f(y))}_{L_{2.42}} \leq 0. \tag{2.42}$$

Necessity of condition 2 is proved by contradiction. Assume (2.42) does not hold for some $y \in Y^f$ with $f(y) = g$, $z(g) \leq g$, $y_i := a$, $y_j := b$ with $0 \leq a < b < z(g)$.[65] Therefore, at (a, g), $(b, g) \in X_p$, we have for some $\ell > 0$ that $L_{2.42} = \ell$. I prove that for all $\epsilon > 0$, there exists $y' \in Y^f$ with $f(y') = g$ such that $|\ell - L_{2.41}(y')| < \epsilon$ and hence there exists an $y' \in Y^f$ such that $L_{2.41}(y') > 0$. Construct y' such that

- $y'_1 := a$,
- $y'_k := b$ for all $2 \leq k \leq xn - 1$,
- $y'_{xn} := xng - a - (xn-2)b$ and
- $y'_j = y'_{xn}$ for all $xn + 1 \leq j \leq n$.

[64] We assume for simplicity that $xn \in \mathbb{N}$.

[65] I take $a < b$ without loss of generality as the same reasoning can be held for the other assumption.

Notice $y'_{xn} > z(g)$ since $a < b < g$ and $g \geq z(g)$, which implies that $y' \in Y^f$. For this y', as $y'_{xn} > z(g)$ we have for all $i \in \{xn, \dots, n\}$ that $\partial_1 d(y'_i, g) = 0$. Therefore

$$\begin{aligned}\ell - L_{2.41}(y') &= L_{2.42} - L_{2.41}(y')\\ &= -\frac{1}{xn}\Big(2\partial_1 d(b,g)\ s(b,g) - \partial_1 d(a,g)\ s(a,g)\Big).\end{aligned}$$

Taking $n(y')$ sufficiently large, we can make $|\ell - L_{2.41}(y')| < \epsilon$, implying $L_{2.41}(y') > 0$, which violates (2.41) and hence modified *Monotonicity in Income* does not hold. The case for which $\partial_1 d(b,g)$ and $\partial_1 d(a,g)$ are not finite is treated as in the proof of Lemma 6.

Sufficiency of condition 2 follows from the fact that if there exists an $y \in Y^f$ violating (2.41), inequality (2.42) is violated as well for a particular value of y_j. For all $y \in Y^f$ there exists $y^*_j \in \big[0, z(f(y))\big)$ such that, taking $y_j := y^*_j$ in $L_{2.42}$, we have $L_{2.41}(y) < L_{2.42}$:

$$-\frac{1}{xn}\sum_{j=1}^{n}\partial_1 d(y_j, f(y))\ s(y_j, f(y)) < -\partial_1 d(y^*_j, f(y))\ s(y^*_j, f(y)),$$

$$-\frac{1}{xn}\sum_{j=1}^{xn}\partial_1 d(y_j, f(y))\ s(y_j, f(y)) < -\partial_1 d(y^*_j, f(y))\ s(y^*_j, f(y))$$

where the strict inequality comes from the presence of the non-poor agent xn. Observe we can consider only the xn first agents since for all agents $k \in \{xn, \dots, n\}$, we have $y_k > z(f(y))$ and hence $\partial_1 d(y_k, f(y)) = 0$.[66] At the value of reference statistic $f(y)$, y^*_j is obtained by solving the following problem:

$$y^*_j := \arg\max_{y_j \in [0, z(f(y)))} -\partial_1 d(y_j, f(y))\ s(y_j, f(y)).$$

■

The necessary condition and the sufficient condition for modified *Monotonicity in Income* are direct modifications of that obtained when mean income is the income standard. The proof of Theorem 10 is not shown as it is done by following the same argument as the one given in the proof of Theorem 2, the equivalent theorem for mean income.

■

2.9.9 Proof of Theorem 7

The proof is by contradiction. Assume that EO $\succeq$ is such that for some $(y^*_i, f^{gm}(y^*)) \in X_p$ we have $k := s(y^*_i, f^{gm}(y^*)) > 0$.

The modified version of Translation Monotonicity imposes that for all $(y_i, f^{gm}(y)) \in X_p$ and all $\delta > 0$ we have

$$(y_i + \delta, f^{gm}(y + \delta\mathbf{1_n})) \succeq (y_i, f^{gm}(y)).$$

[66] If $y_{xn} = z(f(y))$ and $s(f(y)) > 0$, then the increment ϵ given to a poor agent implies $y' \notin Y^f$ as $y'_{xn} < z(f(y'))$.

Take any continuous and differentiable numerical representation d of $\succeq$ (such d exists since $\succeq$ is a continuous ordering). Translation Monotonicity can be equivalently restated using the numerical representation d: for all $y \in Y$, $i \leq q$ and $\delta > 0$ we have:

$$d(y_i + \delta, f^{gm}(y + \delta \mathbf{1_n})) \leq d(y_i, f^{gm}(y)).$$

A necessary condition for the previous condition to hold is that we have for all $y \in Y$ and $i \leq q$[67]

$$\partial_1 d(y_i, f^{gm}(y)) + \partial_2 d(y_j, f^{gm}(y)) \, (\nabla f^{gm}(y) \cdot \mathbf{1_n}) \leq \mathbf{0}.$$

Given the slope is defined as

$$s(y_i, f^{gm}(y)) := -\frac{\partial_2 d(y_j, f^{gm}(y))}{\partial_1 d(y_j, f^{gm}(y))},$$

at $(y_i^*, f^{gm}(y^*))$, the previous inequality amounts successively to:

$$\partial_1 d(y_i^*, f^{gm}(y^*)) - k\partial_1 d(y_i^*, f^{gm}(y^*)) \;\; (\nabla f^{gm}(y^*) \cdot \mathbf{1_n}) \leq 0,$$
$$\partial_1 d(y_i^*, f^{gm}(y^*)) \, (1 - k(\nabla f^{gm}(y^*) \cdot \mathbf{1_n})) \leq 0.$$

Since the first factor is strictly negative by modified Strict Monotonicity in Income, a necessary condition for Translation Monotonicity is that:

$$k \, (\nabla f^{gm}(y^*) \cdot \mathbf{1_n}) \leq 1. \tag{2.43}$$

I construct $y^1 \in Y \subset \mathbb{R}_+^N$ such that $(y_{n-1}^1, f^{gm}(y^1)) = (y_i^*, f^{gm}(y^*))$ and (2.43) is violated at y^1, leading to a violation of Translation Monotonicity. Let y^1 be constructed such that

- $y_j^1 := 0$ for all $j \leq n - 2$,
- $y_{n-1}^1 := y_i^*$, and
- y_n^1 is such that $y_n^1 \geq z(f^{gm}(y^1))$ and $f^{gm}(y^1) = f^{gm}(y^*)$.

For n sufficiently large, there exists such an y_n (unshown). By the definition of the generalized mean, we have:

$$\begin{aligned}\partial_i f^{gm}(y) &= \frac{1}{n} \left(\frac{y_1^\beta + \cdots + y_n^\beta}{n} \right)^{\frac{1}{\beta}-1} y_i^{\beta-1}, \\ &= \frac{1}{n} \big(f^{gm}(y) \big)^{1-\beta} y_i^{\beta-1}.\end{aligned}$$

Therefore, we have:[68]

$$\nabla f^{gm}(y^1) \cdot \mathbf{1_n} \geq \tfrac{n-2}{n} \big(f^{gm}(y^1) \big)^{1-\beta} 0^{\beta-1}.$$

As I assumed $\beta < 1$, the factor $0^{\beta-1} = +\infty$ and (2.43) is violated. As a result, Translation Monotonicity cannot hold.

---

[67] Where $\cdot$ is the notation for a dot product and ∇ is the notation for the gradiant.

[68] The inequality sign comes from the fact I ignored the positive terms coming from the increments given to agents $n - 1$ and n.

2.9.10 Mexican poverty: further analyzis

This section aims at presenting two extra graphical tools. These tools are modifications of well-known tools introduced in the poverty measurement literature, which are both intuitive and helpful in analyzing the evolution of income poverty. I illustrate the changes that occured in Mexico using these graphical tools.

The economic growth of Mexico between 1990 and 2010 has lead to an almost complete eradication of absolute poverty. Nevertheless, the increase in income inequality over that period, as measured by the relative measure HC^{RL}, increased the fraction HC^{EL} of poor individuals. P^{EL} concludes that income poverty has not changed, even if its nature became more relative than absolute.

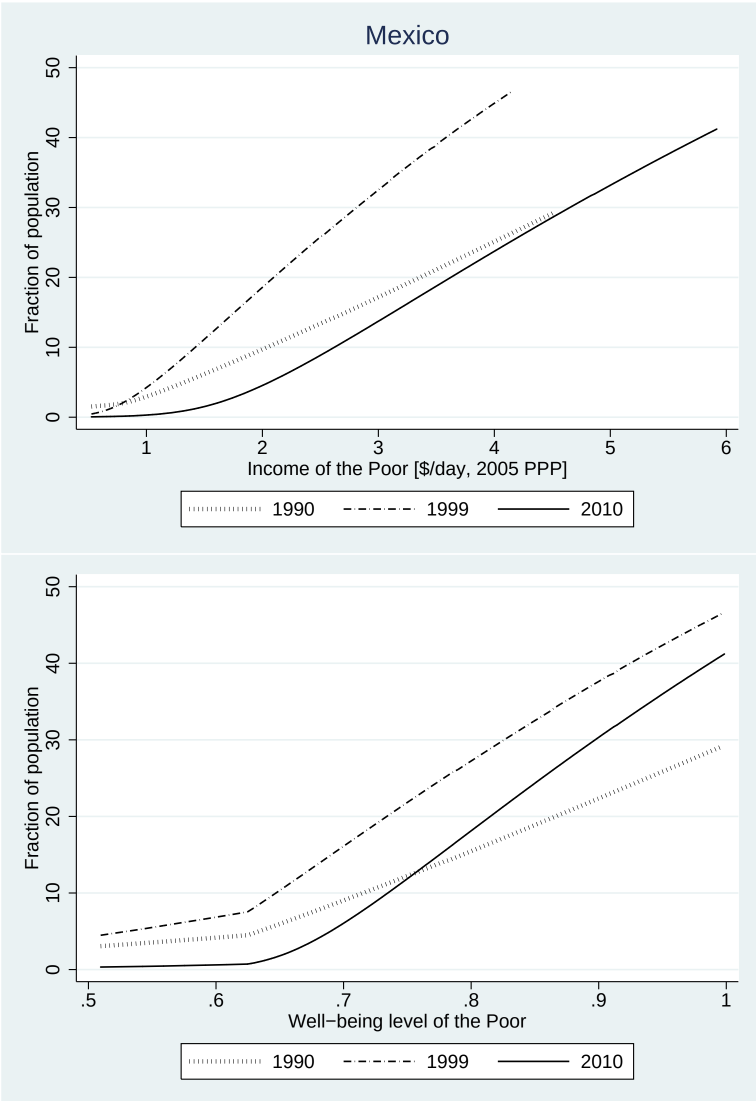

Figure 2.8: Evolution of Mexico's cumulative distributions below the endogenous line. The well-being levels presented above are defined as $1 - d(y_i, \overline{y})$. Source: PovcalNet.

Figure 2.8 shows for several points in time cumulative distributions limited to individuals whose income is below the endogenous threshold. The upper figure shows standard cumulative income distributions. The lower figure shows cumulative well-being distributions. Well-being is defined as $1 - d(y_i, \overline{y})$, where the numerical representation d is based on the absolute-homothetic EO illustrated in Figure 2.6.a. Any individual with a well-being of one or above is non-poor.

The two graphs are such that the cumulative distributions for 2010 first-order stochastically dominate those of 1999. There is hence an unambiguous improvement over that period of both incomes and well-beings. The cumulative income distributions inform on the evolution of several variables. As it is limited to poor individuals, it shows the evolution of the endogenous threshold, which was \$4.5 a day in 1990, \$4.2 a day in 1999 before increasing up to \$5.9 a day in 2010. This evolution translates the changes in mean income, from \$7.8 a day in 1990 to \$7.1 a day in 1999 and up to \$10.6 a day in 2010. The evolution of HC^{EL}, from 29% in 1990 and 47% in 1999 to 41% in 2010, can be read from the graphs by considering the end points' ordinate of these cumulative distributions. Similarly, the same graph presents the evolution of HC^{AL}, from 4.5% in 1990 and 7.5% in 1999 to 0.7% in 2010.

Income cumulative distributions can be drawn without making any normative choice on how to balance absolute and relative income in individual well-being (except those already made by the endogenous line). The cumulative well-being distributions in the lower graph make such choices with its absolute-homothetic EO. These graphs provide again the evolution of HC^{EL}. More interestingly, they show the evolution of P^{EL}. The values of P^{EL} is equal to the areas below the cumulative well-being distributions.[69] Comparing the graphs related to 1990 and 2010 shows that less individuals have very low well-being in 2010 than in 1990 (the well-being threshold for absolute poverty is $\frac{1.25}{2} = 0.625$) but more individuals have well-being levels between 0.75 and 1, indicating that more individuals are in relative poverty. The index balances these two aspects and concludes income poverty has not changed (the area below the two curves is the same): P^{EL} equals 7.4% in 1990, 11.6% in 1999 and 7.5% in 2010.

If the EO used for assessing the well-being of individuals is homothetic, then the area below the cumulative distributions of "homothetic" well-being equals PGR^{EL}. This cumulative distribution of "homothetic" well-being, which gives no priority to individuals below the subsistence level z^a, concludes that income poverty has changed from 12.4% in 1990, 20.1% in 1999 to 15.6% in 2010.

[69] Letting $F : [0,1] \to [0,1] : 1 - d(y_i, \overline{y}) \to F(1 - d(y_i, \overline{y}))$ be the cumulative well-being distribution function associated with $\succeq^{AH}$, we have that $P^{EL} = \int_0^{HC^{EL}} 1 - F^{-1}(x)dx$, where $1 - F^{-1}(x)$ is the individual poverty level such that a fraction x of the population has lower well-being.

Chapter 3

A general criterion to compare mechanisms when solutions are not unique, with applications to school choice

(Joint with Martin Van der Linden)

3.1 Introduction

The literature on mechanism design studies the institutions through which individuals interact in order to reach an economic decision. The individuals hold private information that is relevant to the decision and they are assumed to interact strategically through the institutions. An institution is modeled as a *mechanism* specifying the "rules of the game". A mechanism describes which messages individuals can send and how these messages translate into a decision that defines an economic *outcome*. The aim of this literature is to design mechanisms that satisfy key properties. For instance, a mechanism should incentivize the individuals to truthfully reveal their private information and the messages sent by individuals should lead to fair and efficient outcomes.

When comparing the outcomes of two mechanisms ψ and φ, it is standard to adopt the following two-stage procedure. First, one chooses a solution concept C that *uniquely* predicts the way agents play for each type profile in ψ and in φ (or at least C yields a unique *outcome* for every type profile). Then, one determines whether the unique C-outcomes of ψ and φ satisfy some property of interest, say property X, for *every* possible type profile in a relevant domain. When this is the case for φ but not for ψ, one then concludes that φ is X and ψ is not X. In our terminology, this binary categorization of mechanisms φ and ψ with respect to X implies that φ is *more* X than ψ.

For instance, consider the school choice problem for which a set of school seats has to be allocated among a set of students. Suppose that φ is the Deferred Acceptance mechanism (DA), whereas ψ is the Top Trading Cycle mechanism (TTC). Let C be "dominant strategy", and X be "stability", a fairness property guaranteeing that no student has justified envy given the allocation of seats. The dominant strategy outcome of DA is stable whatever the type profile, whereas the dominant strategy outcome of TTC is not necessarily stable (see for example Haeringer and Klijn (2009)). From this observation, one would typically conclude that DA is more stable than TTC.

As elegant as this standard technique may be, it presents two major limitations that make it applicable to only a handful of mechanism comparisons. Most mechanisms cannot be *uniquely* solved via a convincing solution concept. In particular, comparisons in dominant strategies are known to severely restrict the set of comparable mechanisms. Second, it is often the case that both ψ and φ fail to satisfy property X for *some* (but not all) of the type profiles in the domain. In this case, ψ and φ cannot be compared in terms of X using the standard technique. We refer to these two limitations as the *multiple solutions* issue and the *subdomain violation* issue.

In this paper, we introduce a simple but powerful criterion that generalizes the standard technique and allows us to compare a wide variety of mechanisms. Informally, we will say that a mechanism φ is *at least as X as* ψ if, for each type profile at which all outcomes satisfy X in ψ, all outcomes satisfy X in φ. A formal definition of our criterion can be found in Section 3.2. In that section, we also discuss how our criterion encompasses and extends techniques formerly used in the literature to overcome the limitations of the standard technique.

In the rest of the paper, we illustrate the usefulness of our criterion by applying it to stability comparisons of constrained school choice mechanisms (Haeringer and Klijn, 2009). A school choice mechanism is *constrained* if students can only report preferences over a limited number of schools. Constrained school choice mechanisms are well-suited to showcase the use of our criterion because they combine both of the aforementioned difficulties. In general, constrained school choice mechanisms have no dominant-strategy and cannot be uniquely solved using reasonable solution concepts. Also, for many combinations of constrained school choice mechanisms and solution concepts, the resulting outcome is stable under *some* type profiles only.

In practice, constrained school choice mechanisms are the norm rather than the exception.[1] In a recent paper, Pathak and Sönmez (2013) show that in many mechanisms, increasing the number of schools that students can rank decreases the vulnerability of the mechanism to manipulations by the students. A natural question is whether this decrease in manipulability comes at the cost of a decrease in stability. Given that constained school choice mechanisms are subject to the multiple solutions issue for credible solution concepts, the standard mechanisms comparison techniques can not be used. Using our new criterion, we are able to answer this question.

[1] As observed by Pathak and Sönmez (2013) and others, it is rare that school districts allow students to rank all the schools they could potentially be assigned to. For instance, at the time Haeringer and Klijn (2009) was written, the authors reported that the New York City school district allowed students to rank only 12 programs, while the district had more than 500 different programs available.

We focus on the constrained versions of the so-called *Boston mechanism* (BOS) and *Deferred Acceptance mechanism* (DA). Through our analysis of BOS and DA, we illustrate how our criterion can be applied using different solution concepts, and how the choice of a solution concept can influence our comparisons. Roughly, we show that if students play Nash equilibria (NE), increasing the number of schools that can be ranked actually *decreases* stability in both BOS and DA. Also, under NE, constrained BOS are more stable than constrained DA. However, if students play undominated strategies (US), both conclusions are reversed. That is, increasing the number of schools that can be ranked *increases* stability, and constrained DA are more stable than constrained BOS.

The paper is organized as follows. Our criterion and its relation with other comparison techniques are presented in Section 3.2. The two families of mechanims studied are defined in Section 3.3. The stability comparisons obtained with our criterion are in Sections 3.4, 3.5 and 3.6. Section 3.8 concludes.

3.2 A general criterion for comparing direct mechanisms

In this section, we give a general mechanism design formulation of our criterion. We illustrate the definitions with examples from school choice. The reader unfamiliar with the school choice model is referred to Appendix 3.9.1 where the model is described in details.

There is a finite set of players I and a finite set of outcomes A. A generic player is represented by i. Each player is associated with a type $y_i \in Y_i$, where Y_i is the set of possible types for player i. A list of types for every agent $y := (y_i)_{i \in I}$ is a type profile. The set of possible type profiles is $Y := \Pi_{i \in I} Y_i$.

Example 1 (School choice profile)**.** *The set of players I is composed of a set of students and a set of schools. Students' types are determined by their preferences over schools. A school's type is determined by a priority ranking over students and a capacity which determines the number of students it can accept. The sets Y_i represent possible preferences when i is a student, or possible pairs of priorities and capacities when i is a school. The set of outcomes A contains the assignments of students to schools.* ▲

A *direct mechanism* is a function $\varphi : Y \to A$ associating every type profile with a single *feasible* outcome in A. If players reveal profile $y \in Y$, the outcome under mechanism φ is denoted $\varphi(y)$.

Example 2 (School choice mechanisms)**.** *Examples of direct mechanisms are the aforementioned BOS and DA, which will be formally described in Section 3.3. Together, a school choice profile and a direct mechanism define a preference revelation game called a* game of school choice *(Ergin and Sönmez, 2006). In a game of school choice, the feasible outcomes are the assignments of students to schools in which no school exceeds its capacity.* ▲

In a direct mechanism, given a type profile y, a *solution concept* is a subset $C(y) \subseteq Y$ of type profiles. Profiles $\tilde{y} \notin C(y)$ are those profiles we do not expect players to reveal when their true type profile is y. Conversely, any $\hat{y} \in C(y)$ is interpreted as a profile that could be revealed when players' true type profile is y. The outcomes of φ for all possible $\hat{y} \in C(y)$ will be called the set of *C-outcomes* under φ (e.g. the set of NE or US outcomes) and are denoted by $\varphi\big(C(y)\big)$. Similarly, we will refer to $C(y)$ as the set of *C-profiles* given y (e.g. the set of NE or US profiles).

Example 3 (NE in school choice mechanisms)**.** *In school choice mechanism, it is assumed that the schools always reveal their priorities and capacities truthfully, or equivalently, that these are known to the school district officials (see Kesten (2011) for an analysis of capacity manipulation by schools). Therefore, the "NE" solution concept is defined as usual, except that $NE(y)_s = y_s$ for every school s and every type profile y.* ▲

In general, we are interested in knowing whether, given type profile y, the C-outcomes of a mechanism φ satisfy some property X, such as stability or efficiency. Formally, we consider any property that is a correspondence $X : Y \to A$ specifying the set of outcomes $X(y) \subseteq A$ satisfying it. If the mechanism has a unique C-outcome for y, then either φ satisfies X for y or φ violates X for y. When confronted to the multiple solutions issue, it can be that some C-outcomes of y satisfy X, while others don't. Whether φ satisfies X for y or not is indeterminate until we specify what it exactly means when solutions are non-unique. In this paper, we say that φ satisfies X in C for y when *all* C-outcomes of y satisfy X.

Definition 20 (φ satisfies X in C for y)**.** *The mechanism φ satisfies property X in solution concept C for the type profile y if*

$$\varphi\big(C(y)\big) \subseteq X(y)$$

For instance, Example 2 in Ergin and Sönmez (2006) shows a type profile for which one NE outcome of BOS is constrained efficient (i.e. Pareto efficient among the stable assignment), whereas the other is not. For some type profiles however, all the NE outcomes of BOS will be constrained efficient.[2] If this is the case, we will say that, for this type profile, BOS is constrained inefficient in NE.

As described in the Introduction, we then have the following comparison criteria. A mechanism φ is at least as X as mechanism ψ in solution concept C if for any type profile for which ψ satisfies X in C, φ satisfies X in C.

Definition 21 (At least as X as)**.** *A mechanism φ is* at least as X *as mechanism ψ in solution concept C if*

$$\left\{y \in Y | \psi\big(C(y)\big) \subseteq X(y)\right\} \subseteq \left\{y \in Y | \varphi\big(C(y)\big) \subseteq X(y)\right\}.$$

[2] A trivial example is when every student has the same priority at each and every schools (e.g. student 1 has the first priority in every schools, student 2 the second priority at every school, and so on). Then there is only one NE outcome which is the outcome of a serial dictatorship, and is therefore (constrained) efficient.

The corresponding strict comparison follows naturally.

Definition 22 (More X than)**.** *A mechanism* φ *is* more X *than mechanism* ψ *in* C*, if*

(i) φ *is at least as* X *as* ψ*,*

(ii) there exists a type profile for which φ *satisfies* X *in* C*, but* ψ *does not satisfy* X *in* C*.*

Relation to other comparison techniques in the literature

Let us first consider the standard technique. In the standard technique,

(i) the set of C-outcomes needs to be a singleton for every type profiles, and

(ii) φ is said to be more X than ψ if the C-outcome of φ satisfies X for *all* type profiles, whereas the C-outcome of ψ violates X for *some* type profiles.

Clearly our criterion encompasses the standard technique in the sense that

1) every pair of mechanisms that can be compared using C in the standard technique can also be compared using C with our criteria, and

2) our criterion agrees with any comparison made via the standard technique.

An approach often used in practice to extend the standard technique, when the set of C-outcomes is a singleton but the two mechanisms are faced with the subdomain violation issue, consists in replacing (ii) by

(ii)-bis φ is said to be more X than ψ if for every type profile in which the C-outcome of ψ satisfies X, the C-outcome of φ also satisfies X.

The condition (ii)-bis defines the "profile per profile" approach. For instance, this approach is used, at least implicitly, in Barberà and Gerber (2014) and Dasgupta and Maskin (2008).[3] Again, our criterion encompasses this approach in the sense of 1) and 2).

Finally, our criterion can also be used to formalize arguments from the literature involving multiple equilibria. Ergin and Sönmez (2006) proved that the set of NE outcomes of BOS is the set of stable assignments. As noted by Ergin and Sönmez (2006), this provides an argument in favor of DA because the outcome of DA is

[3] In Barberà (2014), C is "iterated elimination of weakly dominated strategy" and X is the possibility of agenda manipulation in two sequential voting rules. Barberà (2014) conclude that the two voting rules are equally manipulable because the domains of preference profiles on which they are manipulable are identical. In Dasgupta and Maskin (2008), C is "truthfull revelation", and X is a collection of 5 properties. Dasgupta and Maskin (2008) conclude that the Condorcet method is best at satisfying X because for any profile in which *any* other voting rule satisfies X, the Condorcet method also satisfies X.

efficient among the stable assignments when students play dominant strategies.[4] Our comparison criterion provides a useful way to formalize this argument. According to our criterion, if students play a NE in undominated strategies, then DA is at least as efficient as BOS.[5]

Our approach also bears some similarities with the recent literature on mechanisms' degree of manipulability.[6] However, our criterion focuses on the properties of mechanisms' *outcomes*, rather than on the extent to which they are manipulable. This requires us to solve the game associated with the mechanisms of interest, whereas notions of manipulability can often be defined without explicit attention to game theoretic considerations. As a consequence, manipulability comparisons need not be confronted to the multiple solutions issue.

A different criterion for outcome comparison is proposed in Chen and Schonger (2012). Neither Chen and Schonger (2012)'s criterion nor ours' encompasses the other in the sense described above. A comparison with our criterion is provided in Section 3.7.

We now turn to an application of our criterion to stability comparison for constrained versions of BOS and DA, considering NE and US as solution concepts.

3.3 Two classes of competing mechanisms

In this section we describe the two classes of games of school choice that we will be interested in. These classes were identified by Haeringer and Klijn (2009) and correspond to constrained versions of BOS and DA. We first describe the well known *un*constrained BOS.

Step 0: Students rank as many schools as they want and report their ranking.

Step 1: Students apply to the school they reported as their first choice. Every school that receives more applications than its capacity starts rejecting the worst applicants in its priority ranking up to the point where it meets its capacity. All other applicants are *definitively accepted* at the schools they applied to, and capacities are adjusted accordingly.

⋮

Step ℓ : Each student who is not yet assigned applies to the school she reported as her ℓth choice. Every school that receives more *new* applications in step ℓ than

[4] While the NE outcome of BOS might very well be a stable outcome which is Pareto dominated by another stable outcome.

[5] All undominated profiles of DA lead to the unique optimal stable assignment. Because the NE outcomes of BOS are stable, if all NE (in undominated strategies) outcomes of BOS are efficient, the optimal stable assignment is also efficient. It follows that DA is at least as efficient as BOS according to our criterion.

[6] See e.g. Aleskerov and Kurbanov (1999); Parkes et al. (2002); Maus et al. (2007a,b); Day and Milgrom (2008); Dasgupta and Maskin (2010); Erdil and Klemperer (2010); Pathak and Sönmez (2013); Andersson et al. (2014); Arribillaga et al. (2014); Fujinaka and Wakayama (2015)).

its *remaining* capacity starts rejecting the worst *new* applicants in its priority ranking up to the point where it meets its capacity. All other applicants are definitively accepted at the schools they applied to, and capacities are adjusted accordingly.

The algorithm terminates either when all reported preferences have been considered, or when every student is assigned to a school. The constrained versions of BOS, which we will denote by BOS^k, are identical to BOS except that at Step 0, students can rank at most k schools.

We now turn to DA. Again, we first describe the famous *un*constrained DA.

Step 0: Students rank as many schools as they want and report their rankings.

Step 1: Students apply to the school they reported as their first choice. Every school that receives more applications than its capacity *definitively rejects* the worst applicants in its priority ranking up to the point where it meets its capacity. All other applicants are *temporarily* accepted at the schools they applied to (they could be rejected at a later step).

⋮

Step ℓ : Each student who was rejected in step $\ell - 1$ applies to the next school in her reported preferences. Every school considers the new applicants of step ℓ *together with* the students it temporarily accepted. If needed, each school starts rejecting the worst students in its priority ranking up to the point where it meets its capacity. All other applicants are *temporarily accepted* to the schools they applied to (they could be rejected at a later step).

The algorithm terminates either when all reported preferences have been considered, or when every student is assigned to a school. The constrained versions of DA, which we will denote DA^k, are identical to DA except that at Step 0, students can rank at most k schools.

3.4 Comparing DA^k for different values of k

3.4.1 Nash equilibrium

As was shown by (Haeringer and Klijn, 2009, Theorem 5.3), for any k, every NE in DA^k is also a NE in DA^{k+1}. This tells us right away that if all NE outcomes of DA^{k+1} are stable, all NE outcomes of DA^k are also stable. Hence, DA^k is at least as stable as DA^{k+1}.

The converse is not true and DA^k is therefore more stable than DA^{k+1}. In the next example, we provide some intuition for this last result. In the example, t_i is a student and s_j is a school. *In all of our examples, unless stated otherwise, each school always has one seat.* The *revealed* preferences of student i are Q_i and her *true* preferences are P_i. The priorities at school s are F_s. The leftmost panel represents the revealed preferences for DA^3. The boxed schools correspond to the outcome of DA^3 under the revealed profile, and the starred outcome is the most efficient stable assignment. An empty parenthesis ( ) means that the rest of the ranking is arbitrary. A parenthesis containing a crossed-out element ($\not{s}$) means that the rest of the ranking cannot contain s, but is otherwise arbitrary.

Example 4.

$$\begin{array}{llll} Q_1: & \boxed{s_3} & s_1 & (\ ) \\ Q_2: & \boxed{s_1} & s_3 & s_4 \\ Q_3: & (\not{s_1}) & & \\ Q_4: & \boxed{s_4} & s_3 & (\ ) \\ Q_5: & (\not{s_4}) & & \end{array} \qquad \begin{array}{lllll} P_1: & s_3 & s_1^* & (\ ) & \\ P_2: & s_1 & s_2^* & s_3 & s_4 \\ P_3: & s_1 & (\ ) & & \\ P_4: & s_4 & s_3^* & & \\ P_5: & s_4^* & (\ ) & & \end{array} \qquad \begin{array}{lllll} F_{s_1}: & t_1 & t_3 & t_2 & (\ ) \\ F_{s_2}: & t_2 & (\ ) & & \\ F_{s_3}: & t_4 & t_2 & t_1 & (\ ) \\ F_{s_4}: & t_2 & t_5 & t_4 & (\ ) \end{array}$$

The outcome of the revealed profile in Example 4 is unstable and admits two blocking pairs: (t_3, s_1) and (t_5, s_4). Nevertheless, profile Q is a NE since neither t_3 nor t_5 would obtain a better assignment by declaring their blocking school. For instance, consider t_3. If t_3 claimed s_1, it would trigger the following rejection chain *(Kesten, 2010):*

1. *t_3 claims a seat at s_1.*
2. *Because t_3 as higher priority at s_1, t_2 is rejected and applies to s_3.*
3. *Because t_2 has higher priority at s_3 than the incumbent t_1, t_1 is rejected and applies to s_1.*
4. *Because t_1 has higher priority at s_1 than t_3, t_3 is rejected from s_1.*

One can see that any unstable NE requires such a rejection chain for each blocking pair. In particular, there is in Example 4 another rejection chain for the blocking pair (t_5, s_4). An important feature of Example 4 is that t_2 is involved in both of these rejection chains. What is more, it can be shown that for these two rejection chains to co-exist, t_2 must *be able to reveal at least three schools (see Claim 1 of Appendix 3.9.4). Thus this assignment cannot be reproduced as a NE in DA^2. As it turns out, DA^2 is in fact stable (again, see Claim 1 of Appendix 3.9.4), which yields the desired counter-example for DA^2 and DA^3.* ▲

Such an example can be found for every k (see Claim 1 of Appendix 3.9.4), which yields the following proposition.

Proposition 10. *For all $k \in \mathbb{N}$, DA^k is more stable in NE than DA^{k+1}.*

Proposition 10 suggests that when agents have sufficient information on each other's preferences and coordinate on a NE, there might be a stability cost to increasing k. This cost in terms of stability contrasts with the decrease in manipulability identified by Pathak and Sönmez (2013).

Proposition 10 relies heavily on the assumption that students play a NE. Because games of school choice are typically one-shot games in which students have little information on each others' preferences, NE might not be a good approximation of the way students play.[7] This motivates the next subsection, in which we repeat the above analysis while only assuming that students play US.

3.4.2 Undominated strategies

In many respects, US is an interesting solution concept for manipulable games of school choice. With high stakes games such as games of school choice, one may expect students to hire experts to learn about the best strategy to adopt. Now, when students have little information about each others' preferences, experts might only be able to recommend that students avoid dominated strategies (see e.g. Roth and Rothblum (1999)). Therefore, US should be pervasive in practice.

While NE can be viewed as too restrictive, US may seem too loose, in that it allows many revealed profiles to be played for every true profile. However, in the games of school choice we are interested in, US is far from being a vacuous solution concept. In fact, the next example shows that finding a US may be tedious. Doing so requires an accurate knowledge of the priority rankings as well as some non-trivial calculations. As a consequence, it significantly constrains the set of strategies that students can realistically play.

Example 5 (Safe set of schools). *Consider the following profile for DA^3.*

$P_1 : \ (\ )$	$F_{s_1} : \ t_1 \quad t_2 \quad t_4 \quad (\ )$
$P_2 : \ (\ )$	$F_{s_2} : \ t_1 \quad t_4 \quad (\ )$
$P_3 : \ (\ )$	$F_{s_3} : \ t_2 \quad t_4 \quad (\ )$
$P_4 : \ s_1 \quad s_2 \quad s_3 \quad s_4 \quad (\ )$	$F_{s_4} : \ t_4 \quad (\ )$

At first glance, it may seem that for t_4, declaring $Q_4 : s_1 \ s_2 \ s_4$ is undominated. By ranking s_4 (where t_4 has the highest priority), t_4 makes sure that if t_1 and t_2 ended up getting the unique seats at both s_1 and s_2, she would not end up unassigned. But notice that if t_1 and t_2 are assigned to s_1 and s_2, t_2 cannot at the same time be assigned to s_3. Thus for t_4, declaring $Q'_4 : s_1 \ s_2 \ s_3$ dominates Q_4.

We call $\{s_1, s_2, s_3\}$ a safe set *(of schools) for t_4 because by declaring s_1, s_2, and s_3, t_4 is certain to be assigned, whatever the other students declare.* ▲

Interestingly, if students play US, the conclusion of Proposition 10 is reversed, that is DA^{k+1} is more stable in US than DA^k, as stated in Proposition 11. For the most part, US in DA^k ranks schools according to the true preferences of the student (Lemma 13 in Appendix 3.9.2). In particular, when the best $q \leq k$ schools in P_i form a safe set (defined in Appendix 3.9.1), or when t_i only has q acceptable

[7] This is especially true for small k. Suppose we view NE as the remaining strategy profile after a process of iterated deletion of dominated strategies. When k is small, there is typically little or no dominated strategies. Therefore, even at the end of the iterated deletion process, all students will be left with many possible strategies, and the chances that they coordinate on a NE are small.

schools, declaring the q most preferred schools without re-ranking is a dominant strategy (Proposition 18 in Appendix 3.9.5).

As k increases, more and more assigned students have a safe set covering their $q \leq k$ preferred schools, or can declare all their acceptable schools. Therefore, they will play these dominant strategies in any US.[8] As we show in Proposition 19 of Appendix 3.9.5, a student who plays a dominant strategy in DA^k cannot be part of a blocking pair. These observations mean that, as k increases, fewer and fewer students are part of a blocking pair, which provides some intuition as to why DA^{k+1} is at least as stable as DA^k.

The next example illustrates a situation in which the reverse is not true, namely a type profile for which DA^2 is stable in US, while DA^1 is not.

Example 6.

$$\begin{array}{llllllll} Q_1 : \boxed{s_2} & Q'_1 : \boxed{s_1} & P_1 : & s_2 & s_1^* & (\,) & F_{s_1} : & t_1 \;\; t_3 \;\; t_2 \\ Q_2 : \boxed{s_1} & Q'_2 : s_1 & P_2 : & s_1 & s_2^* & (\,) & F_{s_2} : & t_2 \;\; t_1 \;\; (\,) \\ Q_3 : \boxed{s_3} & Q'_3 : \boxed{s_3} & P_3 : & s_1 & s_3^* & (\,) & F_{s_3} : & t_3 \;\; (\,) \end{array}$$

Profiles Q and Q' are both undominated in DA^1, and their outcomes are unstable. Their respective blocking pairs are (t_3, s_1) for Q and (t_2, s_2) for Q'. This example is such that all US outcomes in DA^2 are stable, although students find more than two schools acceptable. For the three students, their two preferred schools form a safe set. Hence there exists a unique US profile in DA^2: the profile in which each student declares her two preferred schools without re-ranking. ▲

Such an example can be found for every k (see Claim 6 in Appendix 3.9.4), which yields the following proposition.

Proposition 11. *For all $k \in \mathbb{N}$, DA^{k+1} is more stable in US than DA^k.*

Proposition 11 suggests that when students have little information and can only resort to undominated strategies, increasing the number of schools that students can declare increases stability (in addition to decreasing manipulability (Pathak and Sönmez, 2013)).

3.5 Comparing BOS^k for different values of k

3.5.1 Nash equilibrium

In NE, the comparison of BOS^k for different values of k is a direct consequence of Haeringer and Klijn (2009).[9] Theorem 6.1 in Haeringer and Klijn (2009) is a straightforward extension to BOS^k for all k of Ergin and Sönmez (2006)'s proof that BOS is stable in NE.

[8] Or an *equivalent* dominant strategy, that is a *dominant strategy* which always yields the same outcome whatever the other students declare (e.g. declaring the safe set first, followed by any set of other schools).

[9] Theorem 6.1 in Haeringer and Klijn (2009).

Proposition 12. *For all $k \in \mathbb{N}$, BOS^k is as stable in NE as BOS^{k+1}.*

Proposition 12 says that when students can coordinate on a NE, the number of schools they are allowed to declare does not affect the stability of BOS^k.

3.5.2 Undominated strategies

When looking at US in BOS^k, we need a minor additional assumption that is common in the literature (e.g. in Pathak and Sönmez (2013)). A set of schools is over-supplied if, the schools in this set as a whole have more quotas than the number of students. An over-supplied set is such that the schools in the set can offer a seat at each and every student. From a strategic point of view, a student who declares an over-supplied set of schools in BOS^k is certain to be assigned, i.e. an over-supplied set of schools is a safe set. We will assume that every *over-supplied* set of schools has more than k schools (see (3.46) in appendix for a formal statement of this condition). *Unless stated otherwise, all our results involving BOS^k rely on this assumption.*

In many cases, this condition is satisfied because no set of schools can accept all potential students. This is the case in many public school districts in the United States. This may be due to the existence of outside options in private schools, or to the segmentation of public high schools into different groups of schools, each with a separate assignment procedure. For instance, Pathak and Sönmez (2013) report that there were over 14,000 applicants in 2009 in the procedure assigning seats at 9 selective public high schools in Chicago, while the 9 schools only had 3,040 seats as a whole.

Even when there happens to be an over-supplied set of schools, this set must still contain no more than k schools for the condition to be violated. As mentioned in the Introduction, k is often much smaller in practice than the total number of schools m. Thus, even when a set of schools is over-supplied, the condition is unlikely to be violated. Furthermore, should the condition be violated, it would most likely be in a district where k is very close to m. In these cases, the impact of increasing k would be rather marginal, and the effect of such a measure can therefore be disregarded for all practical purposes.[10]

As when we compared DA^k for different values of k, the result for BOS^k in NE contrasts with the situation in US (Proposition 13 below). The reason BOS^{k+1} is at least as stable as BOS^k in US is that *under the no over-supplied set assumption*, the US outcomes of BOS^{k+1} are nested in the US outcomes of BOS^k (Proposition 16 in Appendix 3.9.4). That is to say, for every US outcome μ of BOS^{k+1}, there exists an US in BOS^k with the same outcome μ. Therefore, whenever all US of BOS^k yield a stable assignment, it directly follows that all US of BOS^{k+1} also yield a stable assignment.

The next example illustrates a situation in which the converse is not true. In most cases, Q_i is undominated in BOS^k if it contains k acceptable schools, whatever

[10] This being said, whether our results for BOS^k in US still hold when the above condition is violated remains an open question.

the order in which the schools are declared (see Lemma 18 in Appendix 3.9.3 for a more detailed analysis of US in BOS^k). However, in the particular case in which t_i has highest priority at her most preferred school, declaring the preferred school first is a dominant strategy. Based on this special case, the next example describes a type profile for which BOS^3 is stable in US, while BOS^2 is not.

Example 7.

$$
\begin{array}{lll}
Q_1: & \boxed{s_1} & (\ ) \\
Q_2: & \boxed{s_2} & (\ ) \\
Q_3: & \boxed{s_3} & s_1 \\
Q_4: & s_3 & s_1 \\
Q_5: & (\ ) &
\end{array}
\qquad
\begin{array}{llll}
P_1: & s_1^* & (\ ) & \\
P_2: & s_2^* & (\ ) & \\
P_3: & s_1 & s_4^* & s_3 \\
P_4: & s_1 & s_3^* & s_4 \\
P_5: & (\not{s_3}) & (\not{s_4}) &
\end{array}
\qquad
\begin{array}{llll}
F_{s_1}: & t_1 & (\ ) & \\
F_{s_2}: & t_2 & (\ ) & \\
F_{s_3}: & t_1 & t_3 & t_4 \\
F_{s_4}: & t_1 & t_4 & t_3
\end{array}
$$

Profile Q is undominated in BOS^2 and its outcome admits the blocking pair (t_4, s_4). This example is such that all US outcomes in BOS^3 are stable, although several students have (potentially) more than three schools acceptable. Students t_1 and t_2 must declare their preferred school first in any US, as they have the highest priority at their preferred schools. Students t_3 and t_4 only have three acceptable schools and must declare all three in any US of BOS^3. As t_5 does not find s_3 or s_4 acceptable, these schools go to either t_3 or t_4 in any US profile outcome. Observe that any distribution of s_3 and s_4 among t_3 and t_4 result in a stable outcome. ▲

Again, the above example generalizes to all k (see Claim 7 in Appendix 3.9.4), which means we have the following proposition.

Proposition 13. *Assume any* over-supplied *set of schools has more than k schools. Then, for all $k \geq 3$, BOS^{k+1} is more stable in US than BOS^k.*

Proposition 13 parallels Proposition 11. When agents can only resort to undominated strategies, increasing the number of schools students can declare increases the stability of BOS^k.

3.6 Comparing BOS^k and DA^k

3.6.1 Nash equilibrium

Using Theorem 6.1 in Haeringer and Klijn (2009) again, one directly obtains that BOS^k is at least as stable as DA^k. The converse is not true for $k \geq 2$. As noted by Haeringer and Klijn (2009), DA^1 and BOS^1 are formally equivalent.[11] However for $k \geq 2$, there exist unstable NE in DA^k, as shown in Example 4 (see also Haeringer and Klijn (2009)).[12] Therefore we have the following proposition.

Proposition 14. *For all $k \in \mathbb{N}$, BOS^k is at least as stable in NE as DA^k. For all $k \geq 2$, BOS^k is more stable in NE than DA^k.*

[11] Page 1930 in Haeringer and Klijn (2009).
[12] Example 6.2 in Haeringer and Klijn (2009).

Once again, Proposition 14 contrasts with its counterpart in US. As in the comparison of BOS^k with BOS^{k+1}, the reason DA^k is at least as stable as BOS^k in US is that the US outcomes of DA^k are nested in the US outcomes of BOS^k.

3.6.2 Undominated Strategies

Another way to understand proposition 14 is to remember that, as explained in subsection 3.4.2, more and more students play a dominant strategy in DA^k as k increases. As we explain, this implies that fewer and fewer students are part of a blocking pair. It is also the case that students who play a dominant strategy in BOS^k cannot be part of a blocking pair (Proposition 21 in Appendix 3.9.5). But in BOS^k, the only dominant strategies arise when a student has the highest priority at her most preferred school or has only one acceptable school (Proposition 20 in Appendix 3.9.5)). Because this is independent of k, the number of students who play dominant strategies in BOS^k is fixed for all k.

The next example shows that BOS^k is not at least as stable in US as DA^k.

Example 8.

$$\begin{array}{llll} Q_1: & \boxed{s_1} & (\,) & \\ Q_2: & s_1 & s_2 & \\ Q_3: & \boxed{s_2} & s_3 & \end{array} \qquad \begin{array}{llll} P_1: & s_1^* & (\,) & \\ P_2: & s_1 & s_2^* & (\,) \\ P_3: & s_2 & s_3^* & s_1 \end{array} \qquad \begin{array}{lll} F_{s_1}: & t_1 & (\,) \\ F_{s_2}: & t_2 & (\,) \\ F_{s_3}: & t_3 & (\,) \end{array}$$

Profile Q, which is the truncation of students' preferences after their second preferred school, is undominated in both BOS^2 and DA^2. Its outcome in BOS^2 (boxed) is unstable, with blocking pair (t_2, s_2). The fact that t_2 has a higher priority at s_2 than t_3 has been denied by BOS^2 because t_2 did not declare s_2 as her favorite school. Profile Q is the only US profile in DA^2 and its outcome (starred) is the most efficient stable assignment. It is unique as all students have a safe set covering their two preferred schools. ▲

Another source of instabilities in US outcomes of BOS^k that is avoided in DA^k comes from the fact that US of BOS^k may contain non-trivial re-ranking. Although US of DA^k may contain re-ranking, these re-ranking are trivial in the sense that they never influence the outcome (see Lemma 13 in Appendix 3.9.2). On the other hand, US of BOS^k may contain non-trivial re-ranking (like Q_3 in the next example) that turn out to induce instabilities. This is illustrated in the next example.

Example 9.

$$\begin{array}{lll} Q_1: & \boxed{s_1} & (\,) \\ Q_2: & s_1 & s_2 \\ Q_3: & \boxed{s_2} & s_1 \end{array} \qquad \begin{array}{lll} P_1: & s_1^* & (\,) \\ P_2: & s_1 & s_2^* \\ P_3: & s_1 & s_2 \end{array} \qquad \begin{array}{llll} F_{s_1}: & t_1 & (\,) & \\ F_{s_2}: & t_1 & t_2 & t_3 \end{array}$$

Profile Q is undominated in BOS^2 and leads to the boxed unstable outcome, with blocking pair (t_2, s_2). In this example, the unique US profile in DA^2 is P, which leads to the starred stable outcome. ▲

DA^k	$>_{NE}$ $<_{US}$	DA^{k+1}
$\wedge_{NE}{}^{*}$ $\vee_{US}{}^{**}$		$\wedge_{NE}{}^{*}$ $\vee_{US}{}^{**}$
BOS^k	$=_{NE}$ $<_{US}{}^{***}$	BOS^{k+1}

* : $k \geq 2$ ($=_{NE}$ for $k = 1$).
**: no oversupplied set of schools O with $|O| \leq k$.
*** : no oversupplied set of schools with $|O| \leq k$ and $k \geq 3$.

Figure 3.1: Summary of the results. The notation $\varphi >_C \psi$ means "φ is more stable than ψ when students play according to solution concept C".

Constructions like the ones in Examples 8 and 9 can be obtained for all k (see Claim 9 in Appendix 3.9.4), which yields the following proposition.

Proposition 15. *For all $k \in \mathbb{N}$, DA^k is more stable in US than BOS^k.*

Proposition 15 shows that when students cannot coordinate on a NE but rather resort to undominated strategies, DA^k induces less instability than BOS^k.

The results in sections 3.4 to 3.6 are summarized in Figure 3.1.

3.7 On alternative criteria to compare manipulable mechanisms

In this section, we discuss alternative criteria for comparing manipulable mechanisms. We begin with a further consideration of Pathak and Sönmez (2013)'s criterion. Although Pathak and Sönmez (2013)'s criterion is initially defined without reference to a game, they do complement their criterion with a game theoretic interpretation. This interpretation is intimately linked to Pathak and Sönmez (2013)'s concept of a type profiles's *vulnerability to manipulation.*

Definition 23 (Vulnerability to manipulation in mechanism φ (Pathak and Sönmez, 2013))**.** *A type profile y is vulnerable to manipulation under mechanism φ if there exists a player i and a type $y_i' \neq y_i \in Y_i$ such that*

$$\varphi(y_i', y_{-i}) \; P_i \; \varphi(y). \tag{3.1}$$

Notice that (3.1) does not say that truth-telling is a dominant strategy for i. It merely says that for every player i, *whenever everyone else says the truth*, saying the truth is a dominant strategy for i. In other words, a type profile is *not* vulnerable to manipulation in mechanism φ if truth-telling is a NE of φ. Therefore Pathak and Sönmez (2013)'s criteria is equivalent to requiring that everytime *there exists* a truth-telling NE in φ, *there also exists* a truth-telling NE in ψ. One may consider extending this idea to other properties. This would lead to the following criterion.

Definition 24 (Alternative criterion 1)**.** *A mechanism φ is at least as X as mechanism ψ in C if for any type profile for which* there exists *a C-outcome of ψ satisfying X,* there also exists *a C-outcome of φ satisfying X.*

We believe that Pathak and Sönmez (2013)'s criterion was specifically designed for manipulability comparisons. The generalization in Alternative criterion 1 probably goes beyond what they intended. Alternative criterion 1 makes sense when C is NE and X is truth-tellingness. Then, if there exists a truth-telling NE, one may expect that this equilibrium will be played, for many reasons ranging from simplicity, to focal point considerations. In this case, Alternative criterion 1 implicitly says that whenever the *most likely* C-profile of ψ is a truth-telling one, then the *most likely* C-profile of φ should also be truth-telling. This is a very sensible. Things are different with stability or efficiency. It is harder to argue that players will coordinate on efficient equilibria, let alone stable equilibria.[13] Therefore, we believe that Alternative criterion 1 and the criterion we introduce in this paper are complementary. Alternative criterion 1 is best used for manipulability comparisons (as in Pathak and Sönmez (2013)), whereas our criterion is useful in matters of stability and efficiency.

Notice also that the two criteria are not always as different from one another as one may think. For example, the two criteria are identical when the set of C-outcomes tends to a singleton for every type profile y. When the set of C-profiles is a singleton in both φ and ψ, the two criteria boil down to a "problem-by-problem", or "type profile-by-type profile" comparison, as in Barberà and Gerber (2014). In a "type profile-by-type profile" approach, φ is at least as X as ψ if for every type profile y in which the unique C-outcome of ψ satisfies X, the unique C-outcome of φ also satisfies X.

With school choice mechanisms, the set of C-profiles is often a singleton when one takes C to be "dominant strategy". In the criterion of Pathak and Sönmez (2013), letting C be "dominant strategy" is equivalent to replacing Vulnerability to manipulation in mechanism φ by the following vulnerability condition

Definition 25 (Weak vulnerability to manipulation in mechanism φ)**.** *A type profile y is weakly vulnerable to manipulation in mechanism φ if there exists a player i such that truth-telling is not a dominant strategy. That is, for some type $y'_i \neq y_i \in Y_i$* and for some sub-profile of types $y'_{-i} \in Y_{-i}$ (with possibly $y'_{-i} \neq y_{-i}$) ,

$$\varphi(y'_i, y'_{-i}) \; P_i \; \varphi(y_i, y'_{-i}). \tag{3.2}$$

[13] A player can figure out on her own how to play a truth-telling strategy, regardless of what other players declare. On the other hand, one player alone can rarely determine which of her strategies will favor a stable or efficient outcome, as this depends on what the other players declare.

Two other approaches have been suggested by Chen and Schonger (2012). The Theorem 2 in Chen and Schonger (2012) has shown that for the class of mechanism ψ^k studied by these authors, the set of NE leading to a stable outcome in ψ^k is a subset of the set of NE leading to a stable outcome in ψ^{qk}, for any positive integer q. Furthermore these equilibria generate the same outcome in both ψ^k and ψ^{qk}. In general, this suggests the following criterion.

Definition 26 (Alternative criterion 2)**.** *A mechanism φ is at least as X as mechanism ψ in C if for any type profile, the C-profiles of ψ that lead to an outcome satisfying X*

(i) are a subset of the C-profiles of φ, and

(ii) lead to the same outcome in φ (and are therefore a subset of the C-profiles of φ the outcomes of which satisfy X).

A problem with Alternative criterion 2 is that it is silent about the C-profiles which *do not* satisfy X. Suppose that the set of C-profiles satisfying X in ψ is $C^\psi := C^\varphi \cup \{a\}$, for some new element $a \notin C^\varphi$. Suppose also that condition (ii) is satisfied. Then according to Alternative criterion 2, φ is more X than ψ. However, it might very well be the case that at the same time, the set of C-outcomes that do *not* satisfy X in φ is $\bar{C}^\varphi := \bar{C}^\psi \cup \{b_1, \ldots, b_h\}$, for some h arbitrarily large and $b_i \notin \bar{C}^\psi$, for all $i \in \{1, \ldots, h\}$. Then, there is indeed *one* more C-profile in φ than in ψ the outcome of which satisfies X. But at the same time, there are also arbitrarily many new C-profiles in φ the outcomes of which do *not* satisfy X. In this case, one may have some doubts as to whether φ is more X than ψ.

Following the discussion of Alternative criterion 2, a better criterion might be the following, which is illustrated in Figure 3.2.[14]

Definition 27 (Alternative criterion 3)**.** *A mechanism φ is at least as X as mechanism ψ in C if Alternative criterion 2 is satisfied and for every type profile y, the C-profiles of φ that lead to an outcome which does* not *satisfy X*

(i) are also C-profiles of ψ, and

(ii) lead to the same outcome in ψ (and are therefore a subset of the C-profiles of ψ the outcomes of which do not *satisfy X).*

Notice that (i) in Alternative criterion 2 corresponds to the left part of Figure 3.2.

One way to understand Alternative criterion 3 is to assume that, for any mechanism ξ and any solution concept C, each C-profile of ξ is just as likely as any other C-profile of ξ. Then Alternative criterion 3 tells us that obtaining an outcome

[14] The relative size of the sets are irrelevant. Only the inclusion relations are meaningful. Also, the figure only represents the (i) parts of the condition.

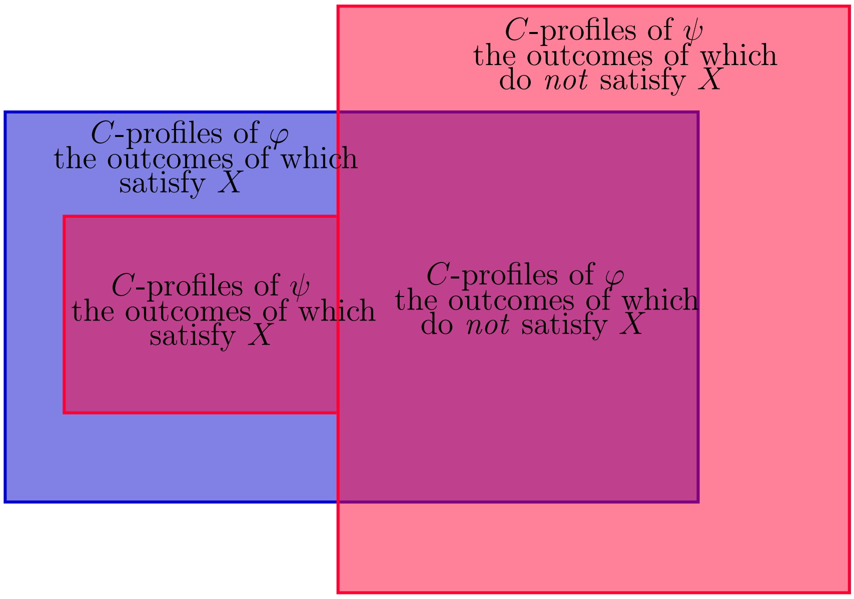

Figure 3.2: Parts (i) in Alternative criterion 3

satisfying X in φ is at least as likely as obtaining an outcome satisfying X in ψ.[15] Under this equal-likelihood assumption, Alternative criterion 3 is of course a very appealing comparison criterion. Without this assumption, things get more complex and there will be cases in which Alternative criterion 3 leads to counter-intuitive comparisons.[16]

Often, even when the above "equal-likelihood" assumption is appropriate, Alternative criterion 3 is unfortunately too demanding to compare mechanisms. For instance, the stability of DA^k cannot be compared using Alternative criterion 3 in US. This can be seen in the following simple example.

Example 10.

$$Q_1 : \boxed{s_1} \qquad P_1 : s_2 \ s_1 \qquad F_{s_1} : t_1 \ (\ )$$
$$Q_2 : \boxed{s_2} \qquad P_2 : s_3 \ s_2 \qquad F_{s_2} : t_2 \ (\ )$$
$$Q_3 : \boxed{s_3} \qquad P_3 : s_1 \ s_3 \qquad F_{s_3} : t_3 \ (\ )$$

[15]Let $c^{\xi}_{X} := |\{C\text{-profiles of } \xi \text{ the outcomes of which satisfy } X\}|$ and $c^{\xi}_{\neg X} := |\{C\text{-profiles of } \xi \text{ the outcomes of which do } \textit{not} \text{ satisfy } X\}|$. Then Alternative criterion 3 tells us that $c^{\varphi}_{X} \geq c^{\psi}_{X}$ and $c^{\varphi}_{\neg X} \leq c^{\psi}_{\neg X}$. One can check that the last inequalities imply $\frac{c^{\varphi}_{X}}{c^{\varphi}_{X}+c^{\varphi}_{\neg X}} \geq \frac{c^{\psi}_{X}}{c^{\psi}_{X}+c^{\psi}_{\neg X}}$. Now, assuming that the probability that any C-profile is played is the same for every C-profile, the last inequality is equivalent to saying that the probability to observe an outcome satisfying X is larger in φ than in ψ.

[16] Using the notation from footnote 15, assume that $c^{\psi}_{X} = 1$, $c^{\psi}_{\neg X} = 1000$, $c^{\varphi}_{X} = 1000$ and $c^{\varphi}_{\neg X} = 1$. According to Alternative criterion 3, φ is clearly more X than ψ. However, suppose that the C-profile of ψ which satisfies X is focal, i.e. it is (much) more likely than other C-profiles. Assume also that the C-profile of φ which does *not* satisfy X is focal. If these are "focal enough", one might want to say that ψ is more X than φ, contrary to what Alternative criterion 3 would conclude.

The depicted revealed profile is a US for DA^1 and yields a stable outcome. Yet, it cannot be a US of DA^2 because in all US of DA^2, students must reveal as many schools as they can. Hence the only US in DA^2 is the true profile which yields a stable outcome but cannot be a US of DA^1 either. Therefore, according to Alternative criterion 3 neither is DA^2 more stable than DA^1, nor is DA^2 more stable than DA^1. That is, the two mechanism are not comparable. This is in spite of the fact that all US in DA^2 are stable whereas some US of DA^1 are not. ▲

Another reason why Alternative criterion 3 could be inapplicable is that it requires the C-profiles to yield the same outcomes in both mechanisms. Although this can work for mechanisms which are "close enough", such as DA^k and DA^{k+1}, it will often fail in general for mechanism which are more fundamentally different, e.g. BOS^k and DA^k.

One way to circumvent the limited applicability of Alternative criterion 3 is to simply look at the fraction of C-profiles the outcomes of which satisfy X in φ and ψ, as in Alternative criterion 4 below. Under the equal-likelihood assumption, Alternative criterion 4 preserves the probabilistic interpretation, while freeing us from the limitation of an approach based on the nestedness of C-profiles ((i) in Alternative criterion 3) and on fixed outcomes ((ii) in Alternative criterion 3).

Definition 28 (Alternative criterion 4)**.** *A mechanism φ is at least as X as mechanism ψ in C if for every type profile y, the* fraction *of C-profiles the outcomes of which satisfy X in φ is higher than the* fraction *of C-profiles the outcomes of which satisfy X in φ. That is, using the notation in footnote 15,*

$$\frac{c_X^{\varphi}}{c_X^{\varphi} + c_{\neg X}^{\varphi}} \geq \frac{c_X^{\psi}}{c_X^{\psi} + c_{\neg X}^{\psi}}. \tag{3.3}$$

Again, the equal-likelihood assumption is essential for Alternative criterion 4 to be relevant. Notice that, although Alternative criterion 4 can usually be used to compare more mechanisms than Alternative criterion 3, condition (3.3) must still hold *for every type profile y*. Often, this is still too demanding for Alternative criterion 4 to be applicable. For instance, we show in Appendix 3.9.5 that Alternative criterion 4 cannot be used to compare the stability in US of DA^2 and BOS^2, whereas our criterion concludes that DA^2 is more stable than BOS^2. Even when the equal-likelihood assumption is appropriate, Alternative criterion 4 and our criterion should therefore be viewed as complementary.

The second approach in Chen and Schonger (2012) consists in assessing properties based on *declared* preferences, rather than on *true* preferences. It presents two great advantages. First, because it does not rely on game theoretic considerations, it allows one to directly use the well-known results on stability and efficiency for non-manipulable mechanisms. One can, for instance, easily show that, in terms of the true preferences, DA^k is stable for any value of k. The second advantage is empirical. Relying on declared preferences allows one to analyze school choice problems using data from actual school districts.

For many properties, however, relying on revealed preferences misses the point. For instance, as argued by Ergin and Sönmez (2006), determining whether a mechanism gives most students their first choice in terms of revealed preferences makes little sense. If every student gets her first declared choice, but declares her third choice first in an attempt to game the mechanism, the mechanism might not be deemed all that efficient.

For similar reasons, one must rely on *true* preferences to capture the *fairness dimension* of stability. Suppose t_2 is assigned to s, t_1 has higher priority than t_2 at some school s, and t_2 would *truly* prefer s to her current assignment. Then we are in a situation of *"true" justified envy*, which many people would deem unfair *per se*. On the other hand, if t_1 only *declared* s above her current assignment, but she *truly* liked her assignment better than s, we would only face a *"declared" unjustified envy*, which most people would deem unproblematic. Clearly, the fairness dimension of stability is a matter of true preferences, and not of revealed preferences.

However, declared preferences can be of a great help with other aspects of stability. The assignments resulting from school choice procedures are regularly challenged in courts on the basis that priorities have not been respected. The risk of an assignment being declared illegal by a court is of primary concern to school choice officials. It is reasonable to expect courts to rule about assignments based on *revealed* preferences, rather than *true* preferences. It is hard to imagine a court ruling in favor of a student who complains about a mechanism because of preferences she did not explicitly state. If a student tries to game the mechanism, the court will most likely hold the student responsible for her attempt at gaming. Therefore, if a mechanism is stable with respect to *revealed* preferences, its assignments should not be invalidated by judges.

Notice that stability with respect to the *true* preferences should also limit legal proceedings against final assignments. One would not expect the assignment to be challenged if the only violation of priorities are with respect to schools that students like *less* than their current assignments. Thus the two approaches are complementary in as much as they capture different aspects of the risk of legal proceedings. Stability with respect to declared preferences determines the risk that courts *rule in favor* of the plaintiff student, while stability with respect to true preferences determines the risk that a student *has an interest* to challenge the assignement in court.

3.8 Conclusion

We have proposed a new criterion for comparing the performance of direct mechanisms. We applied our criterion to games of school choice and obtained stability comparisons which are summarized in Figure 3.1, where US stands for undominated strategy and NE for Nash equilibrium. We believe that our criterion can be fruitfully applied to other properties and mechanisms. In the future, we plan to apply it to efficiency comparison in games of school choice.

Although US may seem more empirically relevant than NE in games of school choice, assignment mechanisms are rarely studied in terms of US. The obvious reason is that US allows for a larger number of potential outcomes. Because comparing many potential outcomes can be cumbersome, NE is usually preferred as a solution concept, as it tends to induce fewer potential outcomes. The criterion we have introduced is one possible answer to this difficulty of comparing multiple potential outcomes. It should therefore facilitate the use of "weaker" solution concepts such as US, whenever these solution concepts are better at describing the way agents play.

As is well known, an NE can be viewed as an iteratively undominated profile in which players correctly anticipate each other's actions Mas-Collel et al. (1995). On the other hand, in US, players do not have enough information on each others' preferences to iteratively eliminate dominated strategies, nor do they correctly anticipate each others' actions. There is of course room for a wide variety of additional information and anticipation structures. In Example 5, we indicated that playing US requires an accurate knowledge of the whole profile of priorities. If one believes that this is still too much information, one may want to consider some solution concept US^{-} in which students play undominated strategies given some *restricted* knowledge of the priorities. Alternatively, one could consider a solution concept US^{+} in which students know the priority structure and also have *some* information about the other students' preferences. This may be useful in mimicking the features of some actual school choice problems. For instance, it is often common knowledge that some schools are highly demanded. To mimic this fact, one may want to assume that students know each others' x best choices, or know the number of students who have some school s among their x preferred school(s). Another interesting middle point between US and NE is the case in which agents know each others' preferences and play iteratively undominated strategies, but do not necessarily anticipate each others' action correctly (i.e. players do not necessarily coordinate on a NE).

Which solution concept should be used to analyze games of school choice depends on which solution concept is best at describing students' behavior. This is eventually an empirical question which would benefit from further investigations in laboratory experiments. Whether any of the alternative solution concepts we just mentioned would yield different stability comparisons than in US and NE is an open question.

Another open question revolves around the applicability of Alternative criterion 4. In many respects, when the equal-likelihood assumption is satisfied (see section 3.7), Alternative criterion 4 is the ideal criterion to compare manipulable mechanisms. However, we have shown that Alternative criterion 4 may be too demanding to be applied (see Example 16). It would be interesting to know whether Alternative criterion 4 is applicable for any mechanisms for some relevant solution concept.

3.9 Appendix

3.9.1 The school choice model, terminology and notation

There is a finite set of schools $S := \{s_1, \ldots, s_m\}$ and a finite set of students $T := \{t_1, \ldots, t_n\}$. Schools are associated with a *priority profile* $F := (F_{s_1}, \ldots, F_{s_m})$ and a *quota profile* $q := (q_{s_1}, \ldots, q_{s_m})$, where $q_s \in \mathbb{N}_+$ denotes the capacity of $s \in S$, i.e. the number of seats available in that school. Students are associated with a *preference profile* $R \coloneqq (R_1, \ldots, R_n)$.

Priorities in F are linear orderings of the students in T, while preferences in $R_i \in R$ are linear orderings of all schools in S *and* t_i herself.[17] A strict preference of t_i for s over s' is denoted by $s\ P_i\ s'$, while $s\ R_i\ s'$ denotes a weak preference, i.e. it allows for $s = s'$.

The higher priority of student t_i over t_j at school s is denoted by $t_i\ F_s\ t_j$. A *assignment profile* is $M \coloneqq (F, q, R)$. The set of all possible preference profiles is $\mathcal{P}$.

The preference profile determines the type of the students (for the simple case in which schools do not manipulate). Students then play a preference revelation game. A typical strategy profile for the game is denoted by $Q \coloneqq (Q_1, \ldots, Q_n)$. Formally, strategies are linear orderings of the schools in S and t_i herself. The fact that t_i declares that she finds s better than s' and hence declares s weakly before s' in Q_i is denoted by $s\ Q_i\ s'$, which also allows for $s = s'$. When necessary, we use H_i to denote a declared profile different from Q_i. There is a **re-ranking** of two different schools s and s' in strategy Q_i if s and s' are declared in Q_i and $s\ Q_i\ s'$ although $s'\ P_i\ s$. Given a strategy Q_i, the **truncation** of Q_i after school s is another strategy Q'_i obtained from Q_i be deleting all schools $s' \in Q_i$ declared after s.

For a given preference profile R, the list containing everyone's preferences but t_i's is denoted by R_{-i}. Similarly, for a given declared profile Q, the list containing everyone's declared preferences but t_i's is denoted by Q_{-i}.

A school $s \in S$ is **acceptable** for t_i if $s\ P_i\ t_i$. If s is acceptable for t_i, we slightly abuse the notation and write $s \in P_i$. We will also write $|P_i|$ for the number of acceptable schools for t_i. Similarly, a school $s \in S$ is **declared** by t_i in Q_i if $s\ Q_i\ t_i$. We again abuse the notation slightly and write $s \in Q_i$ if s is declared in Q_i, and $|Q_i|$ for the number of declared schools in Q_i. By the same token, a subset of schools $S' \subseteq S$ is acceptable or declared for t_i if all the schools in S' are acceptable or declared, which we denote (resp.) $S' \subseteq P_i$ and $S' \subseteq Q_i$.

It will also be useful to identify the x-th ranked school in a preference profile or a declared strategy. If school s is ranked in the x-th position in P_i, we write $s = P_i(x)$. Similarly, if s is declared in the x-th position in Q_i, we write $s = Q_i(x)$.

A mechanism Φ associates every $Q \in \mathcal{Q}$ with some assignment of seats for the students. Let $\Phi_i(Q)$ be the seat assigned to student t_i in mechanism Φ when the students report Q. If $\Phi_i(Q) \neq t_i$, t_i is **assigned** in Φ given Q. On the other hand, t_i is **unassigned** in Φ given Q if $\Phi_i(Q) = t_i$.

For every assignment of types, the space of strategy profiles is $\mathcal{Q} = \mathcal{P}$ itself. Whatever the preference relation of the students, students could in theory pretend they have any other preference relation. Given F and q, a solution concept in game Φ is a function $C^\Phi : \mathcal{P} \to 2^{\mathcal{Q}}$ which associates to every potential type profile P a set of strategy profiles which could be played at a C-equilibrium of the game when P prevails (we do not allow for mixed strategies).

The outcome of a game of school choice Φ is a **feasible** assignment $\mu : T \to S$, a function from the set of students to the set of schools such that no school s

[17] An ordering is a complete, reflexive and transitive binary relation. A linear ordering $\succ$ is an ordering that is antisymmetric, that is, $a \succ b$ and $b \succ a$ implies $a = b$.

is associated with more than q_s students. The school assigned to student t_i in assignment μ is noted μ_i. The notation $\mu_i = t_i$ means that student t_i is assigned to herself, or equivalently, is unassigned.

A feasible assignment μ is **efficient** if there exists no other feasible assignment μ' such that for all $t_i \in T$ we have $\mu'_i \ R_i \ \mu_i$ and for at least one $t_j \in T$ we have $\mu'_j \ P_i \ \mu_j$.

A student-school pair (t_i, s) is **blocking** in assignment μ if either t_i prefers s to μ_i (with possibly $\mu_i = t_i$) and s has empty seats in μ, or t_i prefers s to μ_i and there exists t_j with $\mu_j = s$ and $t_i \ F_s \ t_j$. A assignment μ is **stable** if there exists no blocking pair in μ *and* no student is assigned an unacceptable school. For any assignment profile M, there exists a stable assignment. Furthermore, among the set of stable assignments of M, there is one which is deemed at least as good as any other stable assignment by all students, as shown by Gale and Shapley (1962). This Pareto optimal assignment among the stable assignments is called the **most efficient stable assignment**.

For a given mechanism Φ, a set of schools $SS \subseteq S$ is a **safe set** for t_i if for any Q_i in which SS is declared, and for any Q_{-i}, t_i is at least assigned to the worst school of SS according to Q_i under Φ. Formally, $SS \subset S$ is a safe set if $\Phi_i(Q_i, Q_{-i}) \ Q_i \ w^{SS}_{Q_i}$ for every Q_{-i} and $w^{SS}_{Q_i}$ is the worst school in SS according to Q_i. A school s^* is **accessible** for t_i in mechanism Φ given Q_{-i}, if whenever s^* is declared in Q_i and the other students declare Q_{-i}, t_i is assigned a school at least as good as s^* according to Q_i, that is $\Phi_i(Q_i, Q_{-i}) \ Q_i \ s^*$.

For any mechanism Φ and strategy Q_i, the **possible assignment set** of t_i, denoted by $PAS(Q_i)$, is the set of assignments that mechanism Φ assigns to t_i for some Q_{-i}, that is:

$$PAS(Q_i) = \{x \in S \cup \{t_i\} \mid \Phi_i(Q_i, Q_{-i}) = x \text{ for some } Q_{-i}\}.$$

A set of schools S^* is **over-supplied** if there are enough seats in S^* to host all students, that is $\sum_{s \in S^*} q_s \geq |T|$. For any student t_i and any $S^* \subseteq S$, we denote by $w^{S^*}_i$ the worst school in S^* according to P_i, or simply w^{S^*} when no confusion is possible. Finally we say that school s is **safe if favorite** (SIF) for t_i if t_i is among the q_s-students with highest priority in school s. The final terminology refers to the mechanism BOS^k: if school s is SIF for t_i and $Q_i(1) = s$, then $BOS^k_i(Q_i, Q_{-i}) = s$ for all Q_{-i}.

3.9.2 Some useful lemmas about DA^k

Lemma 12. *For any Q, any k, any $\hat{s}$ and any t_i, suppose that $DA^k_i(Q_i, Q_{-i}) = \hat{s}$. Then for all $s^* \neq \hat{s}$ such that $\hat{s} \ Q_i \ s^*$, there exists Q^*_{-i} such that*

(i) t_i's assignment is unchanged, that is

$$DA^k_i(Q_i, Q^*_{-i}) = \hat{s},$$

and

(ii) s^* is available to t_i, that is for $Q_i^* : s^*$, we have

$$DA_i^k(Q_i^*, Q_{-i}^*) = s^*.$$

Proof. Let B be the set of schools that t_i ranks above $\hat{s}$ in Q_i. These are the schools t_i applied to in the course of DA^k under (Q_i, Q_{-i}), but did not get assigned to. Because t_i was rejected from the schools in B, it must be that, in the vector of assignment $DA^k(Q_i, Q_{-i})$, there is another student assigned to each of the available seats in each of the schools in B. Let the set of these students be denoted A.

Now construct Q_{-i}^* as follows :

- For all $t_j \in A$, let Q_j^* be the strategy in which t_j reveals *only* $DA_j^k(Q_i, Q_{-i})$.
- For all $t_h \in T \backslash \{A \cup \{t_i\}\}$, let Q_h^* be any strategy in which t_h does *not* reveals either s^* or $\hat{s}$.

By construction, for every school $s \in B$, there is at least q_s-students with higher priority at s than t_i who rank s first in Q_{-i}^*. Thus t_i will be rejected of any of these schools in $DA^k(Q_i, Q_{-i}^*)$ too. Therefore $DA_i^k(Q_i, Q_{-i}) = \hat{s}$ implies

$$\hat{s}\ Q_i\ DA^k(Q_i, Q_{-i}^*).$$

By construction again, weakly less students declare $\hat{s}$ in Q_{-i}^* than in Q_{-i}. Therefore, $DA_i^k(Q_i, Q_{-i}) = \hat{s}$ implies

$$DA_i^k(Q_i, Q_{-i}^*)\ Q_i\ \hat{s}.$$

Because Q_i is antisymmetric, the last two displayed relations imply

$$DA_i^k(Q_i, Q_{-i}^*) = \hat{s},$$

which proves (i).

By construction again, no-one applies to s^*.

Thus if $Q_i^* : s^*$, we clearly have

$$DA_i^k(Q_i^*, Q_{-i}^*) = s^*.$$

which proves (ii).

■

Lemma 13 (Equivalent US with no re-ranking)**.** *Assume Q_i is an US of DA^k which features re-rankings, i.e. for some s and s' reported in Q_i, $s'\ P_i\ s$ but $s\ Q_i\ s'$. Then the strategy Q_i' which ranks the same schools as Q_i* but without re-rankings *is such that $DA_i^k(Q_i', Q_{-i}) = DA^k(Q_i, Q_{-i})$ for all Q_{-i}.*

Proof. By Lemma 4.2 in Haeringer and Klijn (2008) we have,

$$DA^k(Q_i', Q_{-i})\ R_i\ DA^k(Q_i, Q_{-i}), \qquad \text{for all } Q_{-i}. \tag{3.4}$$

i.e. Q_i' is *not* weakly dominated by Q_i. Then, if in addition there existed some Q_{-i}^* for which,

$$DA^k(Q_i', Q_{-i}^*)\ P_i\ DA^k(Q_i, Q_{-i}^*),$$

it would mean that Q_i' weakly dominates Q_i. But this would contradicts the assumption that Q_i is undominated. Therefore we must in fact have

$$DA^k(Q_i, Q_{-i})\ R_i\ DA^k(Q_i', Q_{-i}), \tag{3.5}$$

for all Q_{-i}. But because R is a antisymmetric, (3.4) and (3.5) imply

$$DA^k(Q_i', Q_{-i}) \ = \ DA^k(Q_i, Q_{-i}),$$

for all Q_{-i}, the desired result. ■

The following claim parallels Lemma 13 for the case of unacceptable schools.

Lemma 14 (Equivalent US without unacceptable schools)**.** *Assume Q_i is an US of DA^k in which some unacceptable schools are ranked, i.e. for some s reported in Q_i we have $t_i \ P_i \ s$. Then the strategy Q_i'' which is identical to Q_i except that Q_i'' does* not *rank the unacceptable schools in Q_i is such that $DA^k(Q_i'', Q_{-i}) = DA^k(Q_i, Q_{-i})$, for all Q_{-i}.*

Proof. By Lemma 13, it is enough to prove the proposition replacing Q_i by a strategy Q_i' ranking the same schools as in Q_i but without re-ranking.[18] Because DA^m is non-manipulable, we have

$$DA^m(Q_i'', Q_{-i}) \ Q_i'' \ DA^m(\tilde{Q}_i, Q_{-i}), \qquad \text{for all } \tilde{Q}_i \text{ and } Q_{-i}.$$

In particular,

$$DA^m(Q_i'', Q_{-i}) \ Q_i'' \ DA^m(Q_i', Q_{-i}),$$
$$\text{for all } Q_{-i} \text{ with } |Q_j| \leq k \text{ for all } t_j \in T.$$

But because DA^k is equivalent to DA^m if we consider only the profiles Q with $|Q_j^k| \leq k$ for all $t_j \in T$, the last displayed relation implies

$$DA^k(Q_i'', Q_{-i}) \ Q_i'' \ DA^k(Q_i', Q_{-i}), \qquad \text{for all } Q_{-i}.$$

But by construction, Q_i'' is without re-ranking, and therefore, the last displayed relation implies

$$DA^k(Q_i'', Q_{-i}) \ R_i \ DA^k(Q_i', Q_{-i}),$$
$$\text{for all } Q_{-i} \text{ such that } DA^k(Q_i', Q_{-i}) \in Q_i'' \cup \{t_i\}.$$

By construction, every acceptable school of Q_i' is ranked in Q_i''. Therefore, the only cases in which $DA^k(Q_i', Q_{-i}^*) \notin Q_i'' \cup \{t_i\}$ is when $t_i \ P_i \ DA^k(Q_i', Q_{-i}^*)$. But because Q_i'' only ranks acceptable schools, $DA^k(Q_i', Q_{-i}) \ R_i \ t_i$ for all Q_{-i} and therefore, in these cases too, $DA^k(Q_i'', Q_{-i}^*) \ R_i \ DA^k(Q_i', Q_{-i}^*)$.

Thus, we have

$$DA^k(Q_i'', Q_{-i}) \ R_i \ DA^k(Q_i', Q_{-i}), \qquad \text{for all } Q_{-i}.$$

which corresponds to (3.4) in the proof of Lemma 13. The rest of the proof is identical to the proof of Lemma 13. ■

Lemma 15 (Equivalent US with $\min\{k, |P_i|\}$ acceptable schools declared.)**.** *Assume Q_i is an US of DA^k in which less than $\min\{k, |P_i|\}$ acceptable schools are declared. Then there exists a strategy Q_i'' ranking $\min\{k, |P_i|\}$ acceptable schools and such that $DA^k(Q_i'', Q_{-i}) = DA^k(Q_i, Q_{-i})$, for all Q_{-i}.*

[18] Indeed, by construction, if Q_i'' is identical to Q_i' except that it does not rank the unacceptable schools of Q_i', Q_i'' is also identical to Q_i except that it does not rank the unacceptable schools of Q_i'. Also, if $DA^k(Q_i'', Q_{-i}) = DA^k(Q_i', Q_{-i})$, for all Q_{-i}, because $DA^k(Q_i', Q_{-i}) = DA^k(Q_i, Q_{-i})$, for all Q_{-i} by Lemma 13, we would have $DA^k(Q_i'', Q_{-i}) = DA^k(Q_i, Q_{-i})$, for all Q_{-i}, the desired result.

Proof. By Lemma 13 and Lemma 14, it is sufficient to prove that there exists Q_i'' such that

$$DA^k(Q_i'', Q_{-i}) = DA^k(Q_i', Q_{-i}), \text{ for all } Q_{-i},$$

where Q_i' only ranks the acceptable schools of Q_i without re-ranking (as by Lemma 13 and Lemma 14, $DA^k(Q_i', Q_{-i}) = DA^k(Q_i, Q_{-i})$, for all Q_{-i}).

Let Q_i'' be any strategy in which $\min\{k, |P_i|\}$ acceptable schools are declared without re-ranking, *including all the (acceptable) schools declared in Q_i'.*

Because DA^m is non-manipulable, we have

$$DA^m(Q_i'', Q_{-i}) \; Q_i'' \; DA^m(\tilde{Q}_i, Q_{-i}), \qquad \text{for all } \tilde{Q}_i \text{ and } Q_{-i}.$$

In particular,

$$DA^m(Q_i'', Q_{-i}) \; Q_i'' \; DA^m(Q_i', Q_{-i}),$$
$$\text{for all } Q_{-i} \text{ with } |Q_j| \leq k \text{ for all } t_j \in T.$$

But because DA^k is equivalent to DA^m if we consider only the profiles Q with $|Q_j^k| \leq k$ for all $t_j \in T$, the last displayed relation implies

$$DA^k(Q_i'', Q_{-i}) \; Q_i'' \; DA^k(Q_i', Q_{-i}), \qquad \text{for all } Q_{-i}.$$

But by construction, Q_i'' is without re-ranking, and therefore, the last displayed relation implies

$$DA^k(Q_i'', Q_{-i}) \; R_i \; DA^k(Q_i', Q_{-i}),$$
$$\text{for all } Q_{-i} \text{ such that } DA^k(Q_i', Q_{-i}) \in Q_i'' \cup \{t_i\}.$$

By construction, every acceptable school of Q_i' is ranked in Q_i''. Therefore, the only cases in which $DA^k(Q_i', Q_{-i}^*) \notin Q_i'' \cup \{t_i\}$ is when $t_i \; P_i \; DA^k(Q_i', Q_{-i}^*)$. But by construction Q_i' also ranks only acceptable schools. Therefore, this last case cannot occur and we have

$$DA^k(Q_i'', Q_{-i}) \; R_i \; DA^k(Q_i', Q_{-i}), \qquad \text{for all } Q_{-i}.$$

which corresponds to (3.4) in the proof of Lemma 13. The rest of the proof is identical to the proof of Lemma 13. ■

3.9.3 Some useful lemmas about BOS^k

Lemma 16 (Any assignment possible when unsafe)**.** *Take any Q_i, an unsafe strategy of BOS^k. Then for all $\ell \in \{1, \ldots, |Q_i|\}$, there exists Q_{-i}^ℓ such that*

$$BOS^k(Q_i, Q_{-i}^\ell) = Q_i(\ell).$$

Proof. By definition of an unsafe strategy, there exists Q_{-i}^* such that

$$BOS_i^k(Q_i, Q_{-i}^*) = t_i.$$

Now consider the Q_{-i}^{**} in which all t_j assigned in $BOS^k(Q_i, Q_{-i}^*)$ declare

$$Q_j^{**}: \quad BOS_j^k(Q_i, Q_{-i}^*) \quad t_i,$$

and for simplicity, students t_h who are unassigned in $BOS^k(Q_i, Q_{-i}^*)$ declare no schools at all in Q_h^{**}.[19] Clearly, we still have

$$BOS_i^k(Q_i, Q_{-i}^{**}) = t_i,$$

as the same set of students apply to $Q_i(1)$ in the first round (hence t_i is still rejected from $Q_i(1)$ in the first round), and all the seats at all schools are filled in the first round. Now construct Q_{-i}^{ℓ} from Q_{-i}^{**} by changing only the declared profile of students $t_j \neq t_i$ who declare $Q_i(\ell)$, and make those students declare no schools at all.

Then if $\ell = 1$, $BOS_i^k(Q_i, Q_{-i}^1) = Q_i(1)$, as requested. Also, if $\ell > 1$, t_i is still rejected from $Q_i(1)$ in the first round and all seats at all schools by $Q_i(\ell)$ are filled in the first round. Therefore, we clearly have $BOS_i^k(Q_i, Q_{-i}^{\ell}) = Q_i(\ell)$, the desired result.

■

Lemma 17. *If Q_i is US in BOS^k, then $PAS(Q_i)$ contains only acceptable schools for i.*

Proof. In order to derive a contradiction, assume there is $s \in PAS(Q_i)$ that is unacceptable. We construct Q'_i dominating Q_i in BOS^k, contradicting the assumption that Q_i is US. We construct Q'_i step by step:

- Step 1: If $Q_i(1) \in P_i$, then $Q'_i(1) := Q_i(1)$. Else $Q'_i(1) := P_i(1)$.
- Step 2: If $Q_i(2) \in P_i$ and $Q_i(2) \neq Q'_i(1)$, then $Q'_i(2) := Q_i(2)$. Else $Q'_i(2) := P_i(1)$ if $Q'_i(1) \neq P_i(1)$, and $Q'_i(2) := P_i(2)$ otherwise.

  ⋮
- Step ℓ: If $Q_i(\ell) \in P_i$ and $Q_i(\ell)$ is not yet declared in $Q'_i(h)$ for $h < \ell$, then $Q'_i(\ell) := Q_i(\ell)$. Else $Q'_i(\ell)$ is the preferred school according to P_i that is not yet declared in $Q'_i(h)$, for $h < \ell$.

  ⋮
- Last step ℓ^* is the minimal step such that either $\ell^* = |PAS(Q_i)|$ or all acceptable schools are declared in Q'_i.

We now prove by contradiction that Q'_i dominates Q_i in BOS^k. First, we show by contradiction that for all Q_{-i}, we have

$$BOS_i^k(Q'_i, Q_{-i}) \ R_i \ BOS_i^k(Q_i, Q_{-i}).$$

Assume there exists Q_{-i} such that

$$\underbrace{BOS_i^k(Q_i, Q_{-i})}_{:=s^Q} \ P_i \ \underbrace{BOS_i^k(Q'_i, Q_{-i})}_{:=s^{Q'}}. \tag{3.6}$$

This implies s^Q is acceptable as by construction, Q'_i contains no unacceptable schools.

Let r^* be the step of algorithm at which t_i is assigned in $BOS^k(Q_i, Q_{-i})$. Let r' be the rank of school s^Q in strategy Q'_i. As a result, if t_i is not assigned a school

[19] This is for simplicity only. By no means does the argument of the proof require that students be allowed to declare no schools. Other more realistic constructions of Q_{-i}^{**} would also do the job.

before step r' of $BOS_i^k(Q_i', Q_{-i})$, then t_i applies to s^Q at step r'. By construction, t_i declares the acceptable school s^Q weakly before in Q_i' than in Q_i. Therefore $r' \leq r^*$.

Now, since by assumption $BOS_i^k(Q_i, Q_{-i}) = s^Q$, the set of $t_j \neq t_i$ who apply to s^Q before step r^*, together with the set of $t_j \neq t_i$ who apply to s^Q in round r^* and have higher priority than t_i at s^Q, has less than q_{s^Q} students. But then, the set of $t_j \neq t_i$ who apply to s^Q before step $r' < r^*$, together with the set of $t_j \neq t_i$ who apply to s^Q in round r' and have higher priority than t_i at s^Q also has less than q_{s^Q} students. Therefore, t_i is assigned a school in $BOS^k(Q_i', Q_{-i})$ at a step of algorithm $r'' \leq r'$, or in other words

$$BOS_i^k(Q_i', Q_{-i}) \; Q_i' \; s^Q.$$

Now, by construction of Q_i', for all ranks $h \in \{1, \ldots, r'\}$, the school $Q_i'(h)$ is by definition

$$Q_i'(h) \; Q_i' \; s^Q, \tag{3.7}$$

and is such that either

(i) $Q_i'(h) = Q_i(h)$, or

(ii) $Q_i'(h) \; R_i \; s^Q$.

In the construction, (ii) corresponds to the cases in which either $Q_i(h) \notin P_i$, or $Q_i(h) \in P_i$ but $Q_i(h) = Q_i'(\underline{h})$ for some $\underline{h} < h$. In these cases, the construction prescribes to set $Q_i'(h)$ to the most preferred school according to P_i which is not yet declared in $Q_i'(\underline{h})$, for some $\underline{h} < h$. Because we only look at h such that (3.7) holds, s^Q has not yet been declared, and hence, (ii) must hold.

Now, let us compare the effect of declaring Q_i with the effect of declaring Q_i' step by step in BOS^k, for steps $r \leq r'$ (when the $t_j \neq t_i$ declare Q_{-i}). Because $r'' \leq r' \leq r^*$, t_i is rejected from the school she applies to in every step $r < r''$ when declaring Q_i. Thus at each step $r < r''$, either

1. (i) holds and t_i is also rejected at step r when declaring Q_i', or
2. (i) does not hold and (ii) holds, that is
$$Q_i'(h) \; P_i \; s^Q \tag{3.8}$$
Then either
   (a) t_i is rejected from $Q_i'(h)$ at step r, or
   (b) t_i is accepted at $Q_i'(h)$ at step r.

But given (3.8), 2.(b) clearly contradicts (3.6). Thus t_i must be rejected at every step $r < r'$ of BOS^k when declaring Q_i'.

Now, this implies that BOS^k will move on to step r', implying $r'' = r'$. But by (3.7), this means

$$BOS_i^k(Q_i', Q_{-i}) \; = \; s^Q,$$

again contradicting (3.6).

We just showed there exists no Q_{-i} such that (3.6) holds. In order to prove that the constructed Q'_i dominates Q_i, there remains to show that there exists Q^*_{-i} such that

$$BOS^k_i(Q'_i, Q^*_{-i})\ P_i\ BOS^k_i(Q_i, Q^*_{-i}).$$

By the definition of $PAS(Q_i)$, for all school $s \in PAS(Q_i)$, there exists Q^s_{-i} such that

$$BOS^k_i(Q_i, Q^s_{-i}) = s.$$

This is also true for any unacceptable school $s' \in PAS(Q_i)$. By assumption, there exists an unacceptable $s' \in PAS(Q_i)$. Since Q'_i contains only acceptable schools, we have that either

- $BOS^k_i(Q'_i, Q^{s'}_{-i}) \in S$, or
- $BOS^k_i(Q'_i, Q^{s'}_{-i}) = t_i$.

In both cases we have $BOS^k_i(Q'_i, Q^{s'}_{-i})\ P_i\ s'$.

■

Lemma 18. *For all $k \in \mathbb{N}$ for which there is no over-supplied set of schools* (3.46), *Q_i is US in BOS^k if and only if*

Case 1 : *$P_i(1)$ is SIF:*

$$Q_i(1) = P_i(1).$$

Case 2 : *$P_i(1)$ is not SIF:*

- *Either $Q_i(1)$ is favorite acceptable SIF,*
- *or $Q_i(1)$ is not SIF and Q_i contains $min\{k, |P_i|\}$ acceptable schools, one of which is preferred to the favorite acceptable SIF.*

Proof. The proof is case by case.

Case 1 : If $P_i(1)$ is SIF, then declaring this school first guarantees t_i to be assigned to her favorite school, for all possible declarations of the other students, showing sufficiency. Condition $Q_i(1) = P_i(1)$ is necessary as any Q'_i for which $Q'_i(1) = P_i(1)$ is such that school $Q'_i(1)$ is less preferred than $P_i(1)$. It is then easy to construct Q^*_{-i} for which $BOS^k_i(Q'_i, Q^*_{-i}) = Q'_i(1)$ (e.g. $Q^*_j(1) \neq Q'_i(1)$ for all $t_j \neq t_i$), which shows necessity.

Case 2 : We first show *sufficiency* of each condition.

First, if $Q_i(1)$ is the favorite acceptable SIF, then Q_i is a safe strategy by Lemma 20. By the definition of an unsafe strategy, only a safe strategy Q'_i can dominate the safe Q_i that guarantees assignement in the acceptable school $Q_i(1)$. By Lemma 20 , because we assume there is no over-supplied set of schools (3.46), Q'_i is safe if and only if $Q'_i(1)$ is SIF. As by assumption $Q_i(1)$ is the favorite acceptable SIF, strategy Q'_i must have $Q'_i(1) = Q_i(1)$ in order to dominate Q_i. As $Q_i(1)$ is SIF, the two strategies lead to the same assignement for t_i, whatever the strategies declared by other students, and are hence equivalent. Therefore no Q_i' dominates Q_i.

Second, if $Q_i(1)$ is not SIF, then Q_i is unsafe by Lemma 20, because we assume there is no over-supplied set of schools (3.46). By Lemma 21, if Q_i is dominated, then Q_i is dominated by a safe strategy Q_i'. Again, the safe strategy Q_i' must be such that $Q_i'(1)$ is SIF (Lemma 20).

Now, by construction, there exists a school $s' \in Q_i$ that is preferred to the favorite acceptable SIF. This guarantees that Q_i' does not dominate Q_i. Indeed, Lemma 16 shows that for any school s declared in an unsafe strategy, there exists a Q_{-i}^* such that

$$BOS_i^k(Q_i, Q_{-i}^*) = s.$$

In particular, there exists Q_{-i}^{**} such that

$$BOS_i^k(Q_i, Q_{-i}^{**}) = s'.$$

As $Q_i'(1)$ is SIF and by assumption $s'\ P_i\ Q_i'(1)$, strategy Q_i' does not dominate Q_i and hence Q_i is undominated.

We then show by contradiction the *necessity* of these conditions.

First, if $Q_i(1)$ is SIF but not the favorite acceptable SIF, it is clearly dominated by Q_i' for which $Q_i'(1)$ is the favorite acceptable SIF.

Second, consider the case in which $Q_i(1)$ is not SIF. Assume first that Q_i contains no school preferred to the favorite SIF. Then it is again dominated by Q_i' with $Q_i'(1)$ being the favorite SIF. Assume now that Q_i contains less than $min(k, |P_i|)$ acceptable schools. Two cases can arise:

- Q_i contains unacceptable schools. As $Q_i(1)$ is not SIF, all schools in Q_i belong to $PAS(Q_i)$. By Lemma 17, Q_i can not be US.
- Q_i contains no unacceptable schools but less than $min(k, |P_i|)$ acceptable schools. There exists hence an acceptable school s that is not declared in Q_i. Strategy $Q_i' : Q_i\ s$ obtained by attaching s at the end of Q_i can be played in BOS^k and dominates Q_i. By construction of Q_i' we have that if

$$BOS_i^k(Q_i, Q_{-i}) \neq BOS_i^k(Q_i', Q_{-i}),$$

then $BOS_i^k(Q_i, Q_{-i}) = t_i$ and $BOS_i^k(Q_i', Q_{-i}) = s$. Since both strategies Q_i and Q_i' are unsafe, there exists such a Q_{-i} by Lemma 21.

■

3.9.4 Proofs of the propositions

Proof of Proposition 10
(DA^{k+1} less stable than DA^k in NE)

We prove in the text that DA^k is at least as stable as DA^{k+1}. Next, we show that DA^{k+1} is *more* stable as DA^k. The required profile for DA^2 and DA^1 is provided in Example 4. We provide the proof for any k in Claim 1.

Claim 1. *For all k, there exists a assignment profile (given in Example 11) such that all NE outcomes in DA^k are stable whereas some NE outcomes in DA^{k+1} are unstable.*

Example 11. *The generic example is an extended version of Example 4 in the text.*

P_1 :	s_3	s_1^*				
P_2 :	s_1	s_2^*	s_3	s_4	$\dots$	s_{k+2}
P_3 :	s_1					
P_4 :	s_4	s_3^*				
P_5 :	s_4					
P_6 :	s_5	s_4^*				
P_7 :	s_5					
$\vdots$						
P_{2k} :	s_{k+2}	s_{k+1}^*				
P_{2k+1} :	s_{k+2}^*					

F_{s_1} :	t_1	t_3	t_2	
F_{s_2} :	t_2			
F_{s_3} :	t_4	t_2	t_1	
F_{s_4} :	t_6	t_2	t_5	t_4
F_{s_5} :	t_8	t_2	t_7	t_6
$\vdots$				
$F_{s_{k+1}}$:	t_{2k}	t_2	t_{2k-1}	t_{2k-2}
$F_{s_{k+2}}$:	t_2	t_{2k+1}	t_{2k}	

In Example 4, we showed that DA^{k+1} is less stable than DA^k for $k = 2$. The example showing that DA^k is more stable than DA^{k+1} is constructed recursively. From the example showing that DA^{k-1} is more stable than DA^k, we add an extra school s_{k+2} and two extra students t_{2k} and t_{2k+1} :

- *School s_{k+2} is attached at the end of preference P_2,*
- *The priority ordering $F_{s_{k+1}}$ is modified in order to give higher priority to t_{2k} than to t_2,*
- *Finally, $F_{s_{k+2}}$, $P_{t_{2k}}$ and $P_{t_{2k+1}}$ are as shown above.*

▲

Proof. The most efficient stable assignment of Example 11 is starred.

The declared profile given below constitutes a NE in DA^{k+1} and leads to the outcome boxed. The pairs (t_3, s_1), (t_5, s_4), $\dots$ and (t_{2k+1}, s_{k+2}) are blocking in this assignment even if the profile is a NE as there exists a rejection chain (Kesten, 2010) for each of these pairs.

Q_1 :	$\boxed{s_3}$	s_1				
Q_2 :	$\boxed{s_1}$	( )	s_3	s_4	$\dots$	s_{k+2}
Q_3 :	( )					
Q_4 :	$\boxed{s_4}$	s_3				
Q_5 :	( )					
Q_6 :	$\boxed{s_5}$	s_4				
Q_7 :	( )					
$\vdots$						
Q_{2k} :	$\boxed{s_{k+2}}$	s_{k+1}				
Q_{2k+1} :	( )					

We prove that there exists no NE of DA^q with $q \in \{1, \ldots, k\}$ leading to an unstable outcome. As NE of DA^q are nested in NE of DA^{q+1}, we need only prove this for $q = k$. The proof is based on the following claim: for all NE outcome, any student t_i having the highest priority in an acceptable school s must be assigned to a school s' with $s' \; R_i \; s$, and can hence never end up unassigned (proof omitted). By construction, all students t_i with $i \in \{1, 2, 4, 6, \ldots, 2k-1, 2k\}$ are in this situation. Two cases can arise for Q_2:

- Case 1: $s_2 \; Q_2 \; s_{k+2}$ or $s_{k+2} \notin Q_2$.
  As a result, the best reply of student t_{2k+1} is to declare her single acceptable school s_{k+2}. Student t_{2k+1} is assigned to s_{k+2} in any NE outcome as she has highest priority in s_{k+2} after t_2 and t_2 receives either s_2 or a school she declares before s_2. Since t_{2k} has highest priority in a school and can therefore not end up unassigned in NE, she is assigned to s_{k+1}. Applying the same reasoning, we have that t_{2k-2} is assigned to s_k, t_{2k-4} is assigned to s_{k-1}, $\ldots$ until t_4 is assigned to s_3 and t_1 is assigned to s_1. This shows t_2 is assigned to s_2 and the assignment obtained is the most efficient stable assignment.

- Case 2: $s_{k+2} \; Q_2 \; s_2$.
  We show by contradiction that such declaration is never a NE. As t_2 has highest priority in school $P_2(2) = s_2$, student t_2 is assigned to a school she deems at least as desirable as $P_2(2)$ for any NE outcome. There are two subcases:

  - Student t_2 is assigned to s_1 in the NE outcome.

    As t_1 and t_4 must be assigned in any NE outcome, student t_1 is assigned to s_3 and consequently student t_4 is assigned to s_4. This reasoning can be pursued until t_{2k-2} is assigned to s_{k+1} and t_{2k} is assigned to s_{k+2}. The pairs (t_3, s_1), (t_5, s_4), $\ldots$ and (t_{2k+1}, s_{k+2}) are blocking in such assignment. As we assumed this assignment is a NE outcome, there must exist a rejection chain for those blocking pairs. Such rejection chain exists for (t_3, s_1) only if $Q_1 : s_3 \; s_1$ and $Q_2 : s_1 \; s_3$ ( ). The rejection chain exists for (t_5, s_4) only if $Q_4 : s_4 \; s_3$ and $Q_2 : s_1 \; s_3 \; s_4$ ( ). This reasoning is pursued until we conclude that the rejection chain for (t_{2k-1}, s_{k+1}) exists only if $Q_{2k-2} : s_{k+1} \; s_k$ and $Q_2 : s_1 \; s_3 \; s_4 \; \ldots \; s_{k+1}$. Therefore $s_{k+2} \notin Q_2$ because Q_2 can contain at most k schools in DA^k, which implies there is no rejection chain for (t_{2k+1}, s_{k+2}), contradicting the hypothesis this assignment is a NE outcome.

  - Student t_2 is assigned to a school s with $s_2 \; P_2 \; s$ in the NE outcome.

    This can not be a NE since t_2 could profitably deviate by declaring s_2 before s.

■

Proof of Proposition 11
(DA^{k+1} more stable than DA^k in US)

For brevity, we will call $*$-*strategy* a strategy featuring *no* re-ranking, *no* unacceptable schools, an in which $\min\{k, |P_i|\}$ schools are declared. Similarly, a $*$-profile is a profile of $*$-strategies. A US*-profile is a US profile which is also a $*$-profile.

Claim 2. *To prove that DA^{k+1} is at least as stable as DA^k in US, it is sufficient to prove that DA^{k+1} is at least as stable as DA^k in US*.*

Proof. Assume that DA^{k+1} is at least as stable as DA^k in US*. This means that whenever DA^k is stable for all US*, DA^{k+1} is also stable for all US*.

Now assume that DA^k is stable in US, with some stable profiles possibly containing non-$*$-strategies. We need to show that DA^{k+1} is also stable in US. Necessarily, DA^k is stable in US* too, because US^* are a subset of US. Therefore, by assumption, DA^{k+1} is stable in US*.

Then consider any US profile $\bar{Q}^{k+1}$ of DA^{k+1}. If $\bar{Q}^{k+1}$ is a $*$-profile, it is stable because DA^{k+1} is at least as stable as DA^k in US* by assumption. Now assume $\bar{Q}^{k+1}$ is *not* a $*$-profile. By Lemma 13, Lemma 14 and Lemma 15, there exists an US* profile $\bar{H}^{k+1}$ with the same outcome as $\bar{Q}^{k+1}$ in DA^{k+1}. By assumption, because $\bar{H}^{k+1}$ is a US* profile, its outcome is stable. Thus because the outcome of $\bar{Q}^{k+1}$ is the same as the outcome of $\bar{H}^{k+1}$, the outcome of $\bar{Q}^{k+1}$ is stable too, the desired result. ■

We now turn to two claims which will help us in the proof of Claim 5, the key step in the proof of the proposition. The next claim says that if a profile dominates another profile in DA^k, it also dominates the other profile in DA^{k+1}.

Claim 3. *Suppose that the reported profile H_i^k dominates Q_i^k in DA^k, and Q_i^k features no re-rankings. Then H_i^k dominates Q_i^k in DA^{k+1} too.*

Proof. Assume not. Then either

$$DA_i^{k+1}(Q_i^k, \hat{Q}_{-i}^{k+1})\ P_i\ DA_i^{k+1}(H_i^k, \hat{Q}_{-i}^{k+1}), \qquad \text{for some } \hat{Q}_{-i}^{k+1} \tag{3.9}$$

or

$$DA_i^{k+1}(Q_i^k, Q_{-i}^{k+1})\ R_i\ DA_i^{k+1}(H_i^k, Q_{-i}^{k+1}), \qquad \text{for all } Q_{-i}^{k+1}. \tag{3.10}$$

Notice that because H^k dominates Q^k in DA^k, there must exists $\tilde{Q}_{-i}^k$ such that

$$DA_i^k(H_i^k, \tilde{Q}_{-i}^k)\ P_i\ DA_i^k(Q_i^k, \tilde{Q}_{-i}^k), \tag{3.11}$$

holds.

But

$$DA_i^k(Q_i^k, \tilde{Q}_{-i}^k) = DA_i^{k+1}(Q_i^k, \tilde{Q}_{-i}^k),$$

and

$$DA_i^k(H_i^k, \tilde{Q}_{-i}^k) = DA_i^{k+1}(H_i^k, \tilde{Q}_{-i}^k),$$

which together with (3.11) implies

$$DA_i^{k+1}(H_i^k, \tilde{Q}_{-i}^k)\ P_i\ DA_i^{k+1}(Q_i^k, \tilde{Q}_{-i}^k),$$

contradicting (3.10). So (3.9) must hold. In what follows, we show that this leads to a contradiction.

Let us construct Q^*_{-i} as follows. For each student $t_j \neq t_i \in T$, let the new reported preferences be $\hat{Q}_j$ in which only $DA_j^{k+1}(H_i^k, \hat{Q}_{-i}^{k+1})$ is declared. That is in $\hat{Q}_{-i}$, every $t_j \neq t_i \in T$ reports only her assignment in DA^{k+1} under profile $(H_i^k, \hat{Q}_{-i}^{k+1})$.

Given the way we constructed $\hat{Q}_{-i}$, it is clear that

$$DA_i^{k+1}(H_i^k, \hat{Q}_{-i}^{k+1}) = DA_i^k(H_i^k, \hat{Q}_{-i}). \tag{3.12}$$

Also because Q_i^k features no re-ranking

$$DA_i^k(Q_i^k, \hat{Q}_{-i}) \; R_i \; DA_i^{k+1}(Q_i^k, \hat{Q}_{-i}^{k+1}). \tag{3.13}$$

The last relation holds because DA satisfies *individually rational monotonicity* (Kojima and Manea, 2010). Roughly speaking, this means that when students rank less unaccessible schools, everyone is weakly better off. In our case, every $t_j \neq t_i \in T$ report (weakly) less preferences in $\hat{Q}_{-i}$ than in $\hat{Q}_{-i}^{k+1}$. Therefore in $\hat{Q}_{-i}$, (weakly) less students than in $\hat{Q}_{-i}^{k+1}$ apply to every school above $DA_i^{k+1}(Q_i^k, \hat{Q}_{-i}^{k+1})$ in t_i's ranking . Therefore, t_i's assignment can clearly not be a lower school in DA^k than in DA^{k+1} according to Q_i^k. But because Q_i^k and P_i agree on the schools ranked in Q_i^k (by assumption, Q_i^k features no re-ranking), t_i cannot be worse off in DA^k than in DA^{k+1} which corresponds to (3.13).

Combining the two last relations with (3.9), we get

$$DA_i^k(Q_i^k, \hat{Q}_{-i}) \; R_i \; DA_i^{k+1}(Q_i^k, \hat{Q}_{-i}^{k+1}) \; P_i \; DA_i^{k+1}(H_i^k, \hat{Q}_{-i}^{k+1}) \;\; = \;\; DA_i^k(H_i^k, \hat{Q}_{-i}),$$

or in short

$$DA_i^k(Q_i^k, \hat{Q}_{-i}) \; P_i \; DA_i^k(H_i^k, \hat{Q}_{-i}) \tag{3.14}$$

But because, by assumption, H^k dominates Q^k in DA^k, we have

$$DA_i^k(H_i^k, Q_{-i}^k) \; R_i \; DA_i^k(Q_i^k, Q_{-i}^k), \qquad \text{for all } Q_{-i}^k, \tag{3.15}$$

where the last two relations form a contradiction.

■

The next claims says that if some strategy $\bar{Q}_i^k$ dominates a $*$-strategy H_i^k with k declared school, then $\bar{Q}_i^k$ always leads to an assignment at least as good as the worst school declared in H_i^k. This implies that $\bar{Q}_i^k$ is a safe strategy.

Claim 4. *In DA^k, suppose that*

1. *$\bar{Q}_i^k$ dominates H_i^k,*

2. *H_i^k ranks k schools,*

3. *H_i^k is a $*$-strategy.*

Let $\underline{s}_H$ be the worse school listed in H_i^k. Then

$$DA^k(\bar{Q}_i^k, Q_{-i}) \ R_i \ \underline{s}_H, \qquad \textit{for all } Q_{-i}. \tag{3.16}$$

Proof. To derive a contradiction, assume that

$$\underline{s}_H \ P_i \ DA_i^k(\bar{Q}_i^k, \hat{Q}_{-i}), \qquad \text{for some } \hat{Q}_{-i}. \tag{3.17}$$

First notice that by Lemmas 13 and 14, there exists some $*$-strategy Q_i^k that also dominates H_i^k. By the domination condition,

$$\underbrace{DA_i^k(Q_i^k, \dot{Q}_{-i})}_{:=\hat{s}} \ P_i \ DA_i^k(H_i^k, \dot{Q}_{-i}), \qquad \text{for some } \dot{Q}_{-i}. \tag{3.18}$$

Because both Q_i^k and H_i^k feature no re-ranking, this implies $\hat{s} \notin H_i^k$. But because H_i^k ranks k schools, this means that there exists $s^* \in H_i^k$ with $s^* \notin Q_i^k$. By definition of $\underline{s}_H$ we have

$$s^* \ H_i^k \ \underline{s}_H. \tag{3.19}$$

But because H_i^k is a $*$-strategy this implies

$$s^* \ R_i \ \underline{s}_H. \tag{3.20}$$

Combined with (3.17) and the fact that $\bar{Q}_i^k$ and Q_i^k always yield the same outcome for t_i, we get

$$s^* \ P_i \ DA_i^k(Q_i^k, \hat{Q}_{-i}), \qquad \text{for some } \hat{Q}_{-i}. \tag{3.21}$$

Following a similar construction to the one we used in Claim 3, we now show that we can alter $\hat{Q}_{-i}$ in such a way that t_i's assignment is unchanged, but t_i could be assigned to t_i if she declared H_i^k, contradicting the fact that Q_i^k dominates H_i^k.

Let B be the set of schools that t_i ranks above $DA_i^k(Q_i^k, \hat{Q}_{-i})$ in Q_i^k. These are the schools t_i applied to in the course of DA^k under $(Q_i^k, \hat{Q}_{-i})$, but did not get assigned to. Because t_i was rejected from the schools in B, it must be that, in the vector of assignment $DA^k(Q_i^k, \hat{Q}_{-i})$, there is another student assigned to each of the available seats in each of the schools in B. Let the set of these students be denoted A.

Now construct $\tilde{Q}_{-i}$ as follows :

- For all $t_j \in A$, let $\tilde{Q}_j$ be the strategy in which t_j reveals *only* $DA_j^{k+1}(Q_i^k, \hat{Q}_{-i})$.
- For all $t_h \in T \backslash \{A \cup \{t_i\}\}$, let $\tilde{Q}_h$ be the strategy in which t_h reveals *only* $\hat{s}$ for some $\hat{s} \neq s^*$.

By construction, for every school $s \in B$, there is at least q_s-students with higher priority at s than t_i who rank s first in $\tilde{Q}_{-i}$. Thus t_i will be rejected of any of these schools in $DA^k(Q_i^k, \tilde{Q}_{-i})$ too. Therefore

$$DA^k(Q_i^k, \hat{Q}_{-i}) \ R_i \ DA^k(Q_i^k, \tilde{Q}_{-i}).$$

By construction again, no-one applies to s^*. Thus because $s^* \in H_i^k$,

$$DA^k(H_i^k, \tilde{Q}_{-i}) \ Q_i \ s^*. \tag{3.22}$$

But because H_i^k is a $*$-strategy by assumption, it features no re-ranking, and therefore

$$DA^k(H_i^k, \tilde{Q}_{-i})\ R_i\ s^*, \tag{3.23}$$

which combined with (3.21) yields

$$DA^k(H_i^k, \tilde{Q}_{-i})\ P_i\ DA^k(Q_i^k, \tilde{Q}_{-i}), \tag{3.24}$$

contradicting the assumption that Q_i^k dominates H_i^k in DA^k.

■

We can now turn to the proof of Claim 5.

Claim 5. *Let H_i^k be an US* in DA^k. Let $\hat{H}_i^{k-1}$ be obtained from H_i^k as follows.*

If H_i^k contains k declared schools : *Let $\hat{H}_i^{k-1}$be obtained by deleting the* last *school in H_i^k,*

otherwise : *set $\hat{H}_i^{k-1} := H_i^k$.*

Then $\hat{H}_i^{k-1}$ is an US^* in DA^{k-1}. To put it differently, for all k and all US^* of DA^k, there exists an US of DA^{k-1} containing the first* $\min(k-1, |H_i^k|)$ *schools declared in H_i^k.*

Proof. That $\hat{H}_i^{k-1}$ is a $*$-strategy of DA^{k-1} is obvious by construction (given that H_i^k is a $*$-strategy by assumption). Thus we only have to prove that $\hat{H}_i^{k-1}$ is a US of DA^{k-1}.

In order to derive a contradiction, assume that there exists a profile Q_i^{k-1} which dominates $\hat{H}_i^{k-1}$ in DA^{k-1}. Then by definition,

$$DA^{k-1}(Q_i^{k-1}, Q_{-i})\ R_i\ DA^{k-1}(\hat{H}_i^{k-1}, Q_{-i}), \qquad \text{for all } Q_{-i}, \tag{3.25}$$

$$\text{and } DA^{k-1}(Q_i^{k-1}, Q_{-i}^*)\ P_i\ DA^{k-1}(\hat{H}_i^{k-1}, Q_{-i}^*), \qquad \text{for some } Q_{-i}^*. \tag{3.26}$$

Let $\hat{Q}_i^{k-1}$ be the declared profile ranking the same schools as Q_i^{k-1}, but without re-rankings. By (Haeringer and Klijn, 2008, Lemma 4.2),

$$DA_i^k(\hat{Q}_i^{k-1}, Q_{-i})\ R_i\ DA_i^k(Q_i^{k-1}, Q_{-i}), \qquad \text{for all } Q_{-i}. \tag{3.27}$$

Thus,

$$DA^{k-1}(\hat{Q}_i^{k-1}, Q_{-i})\ R_i\ DA^{k-1}(\hat{H}_i^{k-1}, Q_{-i}), \qquad \text{for all } Q_{-i}, \tag{3.28}$$

$$\text{and } DA^{k-1}(\hat{Q}_i^{k-1}, Q_{-i}^*)\ P_i\ DA^{k-1}(\hat{H}_i^{k-1}, Q_{-i}^*), \qquad \text{for some } Q_{-i}^*. \tag{3.29}$$

Notice that by construction, neither $\hat{Q}_i^{k-1}$ nor $\hat{H}_i^{k-1}$ feature re-rankings. Therefore, Claim 3 applies and both (3.28) and (3.29) still hold in DA^k, that is,

$$DA^{k}(\hat{Q}_i^{k-1}, Q_{-i})\ R_i\ DA^{k}(\hat{H}_i^{k-1}, Q_{-i}), \qquad \text{for all } Q_{-i}, \tag{3.30}$$

$$\text{and } DA^{k}(\hat{Q}_i^{k-1}, Q_{-i}^*)\ P_i\ DA^{k}(\hat{H}_i^{k-1}, Q_{-i}^*), \qquad \text{for some } Q_{-i}^*. \tag{3.31}$$

Let w^k be the worst school in H_i^k, that is the school that is potentially removed when going from H_i^k to $\hat{H}_i^{k-1}$. Now from $\hat{Q}_i^{k-1}$, construct $\tilde{Q}_i^k$ as follows.

If $H_i^k \neq \hat{H}_i^{k-1}$ (i.e. if we removed a school from $\hat{H}_i^{k-1}$ to construct H_i^k) :
obtain $\tilde{Q}_i^k$ by adding w^k to $\hat{Q}_i^{k-1}$ respecting the true preference order (i.e. in such a way that $\tilde{Q}_i^k$ features no re-rankings).

otherwise : set $\tilde{Q}_i^k := \hat{Q}_i^{k-1}$.

Because H_i^k is US^*, w^k is an acceptable school. Therefore, because $\tilde{Q}_i^k$ features no re-ranking, we clearly have

$$DA_i^k(\tilde{Q}_i^k, Q_{-i}) \; R_i \; DA_i^k(\hat{Q}_i^{k-1}, Q_{-i}), \qquad \text{for all } Q_{-i}, \tag{3.32}$$

and therefore,

$$DA^k(\tilde{Q}_i^k, Q_{-i}) \; R_i \; DA^k(\hat{H}_i^{k-1}, Q_{-i}), \qquad \text{for all } Q_{-i}, \tag{3.33}$$

$$\text{and } DA^k(\tilde{Q}_i^k, Q_{-i}^*) \; P_i \; DA^k(\hat{H}_i^{k-1}, Q_{-i}^*), \qquad \text{for some } Q_{-i}^*. \tag{3.34}$$

Because H_i^k is undominated in DA^k, either

$$DA^k(H_i^k, Q_{-i}^*) \; P_i \; DA^k(\tilde{Q}_i^k, Q_{-i}^*), \qquad \text{for some } Q_{-i}^*, \tag{3.35}$$

$$\text{or } DA^k(H_i^k, Q_{-i}) \; R_i \; DA^k(\tilde{Q}_i^k, Q_{-i}), \qquad \text{for all } Q_{-i}. \tag{3.36}$$

In what follows, we derive a contradiction in these two cases.

Case 1 : (3.35) holds. If $H_i^k = \hat{H}_i^{k-1}$, then (3.35) directly contradicts (3.33). Thus we must have $H_i^k \neq \hat{H}_i^{k-1}$.

Combining (3.35) with (3.33) we get

$$DA^k(H_i^k, Q_{-i}^*) \; P_i \; DA^k(\hat{H}_i^{k-1}, Q_{-i}^*), \qquad \text{for some } Q_{-i}^*. \tag{3.37}$$

Notice that by construction, because $H_i^k \neq \hat{H}_i^{k-1}$, the only difference between H_i^k and $\hat{H}_i^{k-1}$ is that H_i^k contains w^k in last position. Thus it must be the case that

$$DA^k(H_i^k, Q_{-i}^*) = w^k, \tag{3.38}$$

which combined with (3.35) yields

$$w^k \; R_i \; DA^k(\tilde{Q}_i^k, Q_{-i}^*), \qquad \text{for some } Q_{-i}^*. \tag{3.39}$$

Now notice that $\tilde{Q}_i^k$ satisfies the conditions of Claim 4 with respect to $\hat{H}_i^{k-1}$ in DA^k. Thus, (3.16) holds and if $\hat{w}^{k-1}$ is the last school of $\hat{H}_i^{k-1}$ we have

$$DA_i^k(\tilde{Q}_i^k, Q_{-i}) \; R_i \; \hat{w}^{k-1}, \qquad \text{for all } Q_{-i}, \tag{3.40}$$

Then notice that by construction and because $H_i^k \neq \hat{H}_i^{k-1}$

$$\hat{w}^{k-1} \; P_i \; w^k, \tag{3.41}$$

which combined with (3.40) yields

$$DA_i^k(\tilde{Q}_i^k, Q_{-i}) \; P_i \; w^k, \qquad \text{for all } Q_{-i},$$

contradicting (3.39). Therefore case 1 cannot happen.

Case 2 : (3.36) holds. Notice that because we already ruled out case 1, (3.36) is equivalent to

$$DA^k(H_i^k, Q_{-i}) \; = \; DA^k(\tilde{Q}_i^k, Q_{-i}), \qquad \text{for all } Q_{-i}. \tag{3.42}$$

There are two subcases.

Subcase 1 :$DA^k(H_i^k, Q_{-i}) \neq w^k$ for all Q_{-i}.

Then,

$$\begin{aligned} &DA^k(H_i^k, Q_{-i}^{k-1}) \neq w^k, \\ &\quad \text{for all } Q_{-i}^{k-1} \text{ with } |Q_j^{k-1}| \leq k-1 \text{ for all } t_j \neq t_i \in T. \end{aligned} \tag{3.43}$$

This means that for any Q_{-i}^{k-1} with $|Q_j^{k-1}| \leq k-1$ for all $t_j \neq t_i \in T$, t_i is either assigned a school better than w^k or is unassigned, when declaring H_i^k.

In either cases, removing w^k from H_i^k would have no impact on t_i's assignment, regardless of Q_{-i}^{k-1}. Thus we have,

$$\begin{aligned} &DA^k(H_i^k, Q_{-i}^{k-1}) = DA^k(\hat{H}_i^{k-1}, Q_{-i}^{k-1}), \\ &\quad \text{for all } Q_{-i}^{k-1} \text{ with } |Q_j^{k-1}| \leq k-1 \text{ for all } t_j \neq t_i \in T. \end{aligned} \tag{3.44}$$

Because H_i^k is undominated in DA^k, it is undominated for every Q_{-i}^{k-1}. Therefore (3.44) tells us that $\hat{H}_i^{k-1}$ is also undominated in DA^k for all Q_{-i}^{k-1}. In particular, it is not dominated by Q_i^{k-1}. But DA^{k-1} is strategically equivalent to DA^k constrained to the set of profiles with no more than $k-1$ declared schools. Thus (3.44) implies that $\hat{H}_i^{k-1}$ is undominated in DA^{k-1}, a contradiction.

Subcase 2 : $DA^k(H_i^k, Q_{-i}^*) = w^k$ for some Q_{-i}^*. By (3.42), this implies

$$DA_i^k(\tilde{Q}_i^k, Q_{-i}^*) = w^k. \tag{3.45}$$

Then we can use the same kind of constructions as in Claim 4 to construct a profile Q_{-i}^{**} of DA^k such that $\hat{H}_i^{k-1}$ yields a strictly better outcome than $\tilde{Q}_i^k$ in DA^k, contradicting (3.33).

■

With these two lemmas, we are now equipped to prove the "at least as stable" part of the main proposition.

Proof: By Claim 5, for any US* profile Q^{k+1} in DA^{k+1}, there exists an US* profile Q^k in DA^k constructed by removing the worst school of every agent ranking $k+1$ schools in Q^{k+1}. Let μ^k the assignment obtained from Q^k in DA^k. By assumption, μ^k is stable.

For any student t_i, let w_i^{k+1} be the last school t_i declares in Q_i^{k+1}. Now, let T^A be the set of assigned students in μ^k. Because μ^k is stable, and w_i^{k+1} is acceptable (Q_i^{k+1} is an US^*) for every $t_j \in T \backslash T^A$ (the set of *un*assigned students in μ^k), all the seats at every w_j^{k+1} are assigned in μ^k, and they are assigned to students with a higher priority at w_j^{k+1} than t_j. Thus the profile Q_*^{k+1} of DA^{k+1} constructed from Q^k by adding w_i^{k+1} to the declared profile of every $t_i \in T \backslash T^A$ yields the same assignment, that is $DA^{k+1}(Q_*^{k+1}) = \mu^k$.

Now construct a last profile Q_{**}^{k+1} from Q_*^{k+1} by adding w_h^{k+1} to the declared strategy of every $t_h \in T^A$ for which $w_h^{k+1} \notin Q_h^k$. By construction, w_h^{k+1} is below w_h^k, the last school t_h ranks in Q_h^k. Because $t_h \in T^A$, this means t_h is assigned to a better school than w_h^{k+1} in μ^k. Again, this implies that Q_{**}^{k+1} yields the same assignment μ^k than Q_*^{k+1}.

But notice that in constructing Q_{**}^{k+1}, we have added "back" w_i^{k+1} to the declared profile of every agent t_i for whom Q_i^k was constructed from Q_i^{k+1} by removing w_i^{k+1}. Hence, $Q_{**}^{k+1} = Q^{k+1}$, and we have shown that $DA^{k+1} = \mu^k$. Because μ_k is stable by assumption, we have shown the "at least as stable" part of the proposition for US^* profiles. But because of Lemma 2, proving the result for all US^* is enough, and we are done with the proof of the "at least as stable" part of the proposition.

■

Next, we show that DA^{k+1} is *more* stable than DA^k. The required profile for DA^2 and DA^1 is provided in Example 6. We provide the proof for any k in Claim 6.

Claim 6. *For all k, there exists a profile (given in Example 12) such that all US outcomes are stable in DA^{k+1} whereas some US outcomes are unstable in DA^k.*

Proof. The generic example is the following:

Example 12.

$$\begin{array}{llllll} P_1: & s_1^* & (\,) & & & \\ P_2: & s_2^* & (\,) & & & \\ \vdots & & & & & \\ P_{k+1}: & s_1 & s_2 & \dots & s_{k+1}^* & (\,) \\ P_{k+2}: & s_{k+2}^* & (\,) & & & \end{array} \qquad \begin{array}{lll} F_{s_1}: & t_1 & (\,) \\ F_{s_2}: & t_2 & (\,) \\ \vdots & & \\ F_{s_{k+1}}: & t_{k+1} & (\,) \\ F_{s_{k+2}}: & t_{k+2} & (\,) \end{array}$$

▲

All students except t_{k+1} have highest priority in their preferred school. Student t_{k+1} finds k schools better than s_{k+1}, the school she is assigned to in the unique stable assignment (starred). Mechanism DA^k is unstable in US because it is an undominated strategy for t_{k+1} to declare her k preferred schools (and hence not declare s_{k+1}). For any US profile in DA^k declared by the other students, if t_{k+1} does not declare s_{k+1} then the outcome is unstable because t_{k+1} is unassigned whereas s_{k+1} has an available seat. On the other hand, mechanism DA^{k+1} is stable in US since t_{k+1} has a safe set covering her $k+1$ preferred schools, implying her only undominated strategy is a truncation of P_i after s_{k+1}. The unique US profile in DA^{k+1} lead to the unique stable assignment. ■

Proof of Proposition 13
(BOS^{k+1} more stable than BOS^k in US)

This proof of nestedness requires an additional definition. For any strategy Q_i, let $Q_i^{(r)}$ be the strategy obtained from Q_i by deleting the school declared at rank r in Q_i. Formally :

- $Q_i^{(r)}(r') := Q_i(r')$ for all $r' \in \{1, \dots, r-1\}$ and
- $Q_i^{(r)}(r') := Q_i(r'+1)$ for all $r' \in \{r, \dots, |Q_i|-1\}$.

Lemma 19. *Let $\mu := BOS^{k+1}(Q)$ with Q an US profile in BOS^{k+1} and $k \geq 3$. For all $t_i \in T$, if $|Q_i| = k+1$, then there exists $r \in \{|Q_i|-2, |Q_i|-1, |Q_i|\}$ such that $Q_i(r) \neq \mu_i$ and $Q_i^{(r)}$ is US in BOS^k.*

Proof. Take any $t_i \in T$. The proof is case by case.

Case 1 : $Q_i(1)$ is SIF.

By Lemma 18, as Q_i is US in BOS^{k+1}, $Q_i(1)$ is the favorite acceptable SIF. If $r = |Q_i|$, then by Lemma 18 we have $Q_i^{(r)}$ is US in BOS^k as $Q_i^{(r)}(1)$ is the favorite acceptable SIF. Furthermore, $Q_i(r) \neq \mu_i$ as $Q_i(1)$ is SIF, hence $Q_i(1) = \mu_i$, and as we assumed $k \geq 3$, we have $2 \leq |Q_i| - 2 \leq r$.

Case 2 : $Q_i(1)$ is not SIF.

Case 2.1 : $\mu_i = t_i$.

By the case 2 of Lemma 18, as Q_i is US in BOS^{k+1}, all schools declared in Q_i are acceptable and at least one school declared is preferred to the favorite acceptable SIF. Let s^* be the favorite school declared in Q_i. Let $r := |Q_i|$ if $Q_i(|Q_i|) \neq s^*$ and $r := |Q_i| - 1$ otherwise. By construction, as we assumed $|Q_i| = k+1$, strategy $Q_i^{(r)}$ contains k acceptable schools, among which s^* that is preferred to the favorite acceptable SIF. By case 2 of Lemma 18, $Q_i^{(r)}$ is US in BOS^k.

Case 2.2 : $\mu_i \neq t_i$.

The reasoning is the same as for the case above. The only difference is in the construction of r. Rank r is any element in $\{|Q_i|-2, |Q_i|-1, |Q_i|\}$ such that $Q_i(r) \neq \mu_i$ and $Q_i(r) \neq s^*$ where s^* is the favorite school declared in Q_i.

■

Proposition 16. *For all $k \geq 3$ we have $US(BOS^{k+1}) \subseteq US(BOS^k)$.*

Proof: Take any $k \geq 3$ and any profile Q of US in BOS^{k+1}. Let $\mu := BOS^{k+1}(Q)$ be the assignment obtained for the US profile Q. We construct a profile Q' that is US in BOS^k and show that $\mu = BOS^k(Q')$.

Step 1: Construction of profile Q', a US in BOS^k.

We first introduce a particular transformation of a strategy defined for any Q_i and any school $s \in Q_i$. Transformation Q_i^{s-2} of strategy Q_i exchange the ranks of schools $Q_i(2)$ and s. Formally, let r^* be the rank of school s in Q_i:

- If $r^* = 1$ or $r^* = 2$ then $Q_i^{s-2} := Q_i$,
- Else $Q_i^{s-2}(2) := s$, $Q_i^{s-2}(r^*) := Q_i(2)$ and for all $r \in \{1, \ldots, |Q_i|\} \backslash \{2, r^*\}$ we have $Q_i^{s-2}(r) := Q_i(r)$.

For any $t_i \in T$, the construction of Q'_i is case by case. In a nutshell, the first school declared in Q'_i is $Q_i(1)$ and, if the agent is assigned a school in $BOS^{k+1}(Q)$ different from $Q_i(1)$, then the second school declared in Q'_i is μ_i, otherwise we have $Q'_i(2) = Q_i(2)$. The detailed construction of Q'_i goes as follows:

Case 1 : $Q_i(1)$ is SIF.

Take $Q'_i : Q_i(1)$. By Lemma 18, as Q_i is US in BOS^{k+1}, $Q_i(1)$ is the favorite acceptable SIF. Strategy $Q'_i : Q_i(1)$ is US in BOS^k by Lemma 18 as $Q'_i(1)$ is the favorite acceptable SIF.

Case 2 : $Q_i(1)$ is not SIF.

Case 2.1 : $|Q_i| < k + 1$.

By the case 2 of Lemma 18, if Q_i with $Q_i(1)$ is not SIF and $|Q_i| < k + 1$ is US in BOS^{k+1}, we have that

- $|Q_i| = |P_i|$ and strategy Q_i contains all acceptable schools, and
- $P_i(1)$ is not SIF.

By the case 2 of Lemma 18, Q_i is therefore an US in BOS^k. Let $Q'_i := Q_i$ if $\mu_i = t_i$, and $Q'_i := Q_i^{2-\mu_i}$ otherwise. In the latter case, strategy Q'_i is US in BOS^k by Lemma 18 because $Q_i^{2-\mu_i}(1)$ is not SIF, and $Q_i^{2-\mu_i}$ contains all acceptable schools.

Case 2.2 : $|Q_i| = k + 1$.

By Lemma 19, there exists $r \in \{|Q_i|-2, |Q_i|-1, |Q_i|\}$ such that $Q_i(r) \neq \mu_i$ and $Q_i^{(r)}$ is US in BOS^k. For this value of r, let $H_i := Q_i^{(r)}$, strategy H_i is hence US in BOS^k. Let $Q'_i := H_i$ if $\mu_i = t_i$, and $Q'_i := Q_i^{2-\mu_i}$ otherwise. In the latter case, strategy Q'_i is US in BOS^k by Lemma 18 because $H_i^{2-\mu_i}(1)$ is not SIF and $H_i^{2-\mu_i}$ contains k acceptable schools, among which one is preferred to the favorite acceptable SIF. Indeed, the school deleted from Q_i was not the favorite school declared by construction.

Step 2: Proof that $BOS^k(Q') = \mu$.

By the construction of Q', we have $Q'_i(1) = Q_i(1)$ for all $t_i \in T$. As a result, the first step of mechanism BOS^k for the profile Q' is the same as the first step of mechanism BOS^{k+1} for the profile Q. Therefore, from now on, we can focus exclusively on students who are not assigned in the first step of BOS^k.

In the second step of mechanism BOS^k when profile Q' is declared, all students for whom $\mu_i \in S$ declare μ_i and all students for whom $\mu_i = t_i$ declare $Q_i(2)$.[20] We show that all students declaring μ_i in step 2 of BOS^k when Q' is the declared profile are assigned to μ_i and all students for whom $\mu_i = t_i$ are rejected.

[20]When $k = 3$, the construction of Q'_i proposed above does not guarantee that $Q'_i(2) = Q_i(2)$ for students for whom $\mu_i = t_i$. Nevertheless students for whom $\mu_i = t_i$ have an undominated strategy Q_i^* such that both $Q_i^*(1) = Q_i(1)$ and $Q_i^*(2) = Q_i(2)$. Strategy Q_i^* can be constructed in this way even if $k = 3$ by taking either $Q_i^* := Q_i^{(3)}$ or $Q_i^* := Q_i^{(4)}$, depending on the rank of the preferred school declared in Q_i (see case 2.1 in Lemma 19).

- Consider first any student t_i for whom $\mu_i = t_i$.

  As $\mu_i = t_i$, this student is rejected in the second step of mechanism BOS^{k+1} for profile Q. This implies that at least $q_{Q_i(2)}$ other students with higher priority at $Q_i(2)$ than t_i apply at $Q_i(2)$ in the two first step of mechanism BOS^{k+1} for profile Q. By construction, these $q_{Q_i(2)}$ other students with higher priority also apply to $Q_i(2)$ in the two first steps of mechanism BOS^k for profile Q' and none among them is assigned another school in step 1 as the first step is unchanged. As a consequence, t_i is rejected from $Q_i(2)$.

- Consider now any student t_i for whom $\mu_i \in S$.

  By construction we have $Q'_i(2) = \mu_i$ and two cases can be considered. Either there are at most q_{μ_i} students applying to school μ_i during the 2 first steps of mechanism BOS^k for profile Q' and t_i is again assigned to μ_i, or there are more than q_{μ_i} such students. The latter case happens only if all students assigned to μ_i where assigned during the two first steps of mechanism BOS^{k+1} for profile Q. This implies that t_i is among the q_{μ_i} students with highest priority among those who apply to this school during the two first steps of mechanism BOS^{k+1} for profile Q. Student t_i is by construction still among the q_{μ_i} students with highest priority among those who apply to this school during the two first steps of mechanism BOS^k for profile Q'. Therefore t_i is assigned to μ_i.

In later steps of mechanism BOS^k for the profile Q', no more assignements to schools take place. The students remaining unassigned after step 2 are those for whom $\mu_i = t_i$. They apply in later steps to schools from which they were rejected by mechanism BOS^{k+1} for the profile Q, implying all these schools have accepted a number of students equal to their quota. After step 2 of mechanism BOS^k for the profile Q', these schools are also full and therefore student t_i for whom $\mu_i = t_i$ ends up unassigned: $BOS^k_i(Q') = t_i$. This shows that $BOS^k(Q') = BOS^{k+1}(Q)$.

■

We showed that for any $k \geq 3$, any US outcomes in BOS^{k+1} is also an US outcome in BOS^k. As a result, if all US outcomes in BOS^k are stable, all US outcomes in BOS^k are stable as well. This implies that BOS^{k+1} is at least as stable as BOS^k.

Next, we show that BOS^{k+1} is more stable than BOS^k. Example 7 in the text provides the required profile for the comparison of BOS^2 and BOS^3. The generic profile required is given in Claim 7.

Claim 7. *For all k, there exists a profile (given in Example 13) such that all US outcomes are stable in BOS^{k+1} whereas some US outcomes are unstable in BOS^k.*

Proof. The generic example is the following:

Example 13.

P_1 :	s_1^*	( )			
P_2 :	s_2^*	( )			
$\vdots$					
P_k :	s_k^*	( )			
P_{k+1} :	s_1	$\ldots$	s_{k-1}	s_{k+2}^*	s_{k+1}
P_{k+2} :	s_1	$\ldots$	s_{k-1}	s_{k+1}^*	s_{k+2}
P_{k+3} :	s_1	$\ldots$	s_k		

F_{s_1} :	t_1	( )
F_{s_2} :	t_2	( )
$\vdots$		
F_{s_k} :	t_k	( )
$F_{s_{k+1}}$:	t_{k+1}	( )
$F_{s_{k+2}}$:	t_{k+2}	( )

▲

All students except t_{k+1}, t_{k+2} and t_{k+3} have highest priority in their preferred school. By Lemma 18, all students except t_{k+1}, t_{k+2} and t_{k+3} must declare their preferred school first in all undominated strategy. Student t_{k+1} and t_{k+2} find $k-1$ schools better than the school they are assigned to in the most efficient stable assignment μ^e (starred). The only stable assignment different from μ^e is obtained from μ^e by letting t_{k+1} and t_{k+2} exchange s_{k+1} and s_{k+2}. Mechanism BOS^k is unstable in US as there exists Q_{k+1} and Q_{k+2} undominated in BOS^k such that $s_{k+1} \notin Q_{k+2}$ and $s_{k+1} \notin Q_{k+1}$ (see case 2 of Lemma 18). All US profiles for which neither t_{k+1} nor t_{k+2} declare s_{k+1} lead to unstable outcomes: one of these two students is unassigned and the acceptable school s_{k+1} has an empty seat. On the other hand, mechanism BOS^{k+1} is stable since t_{k+1} and t_{k+2} must declare both s_{k+1} and s_{k+2} in all their undominated strategies (they find exactly $k+1$ schools acceptable). ■

Proof of Proposition 14
(BOS^k more stable than DA^k in NE)

We proved in the text that DA^k is at least as stable as BOS^k. Next, we show that DA^k is more stable than BOS^k. Haeringer and Klijn (2008) provide a profile for which DA^2 admits a NE leading to an unstable outcome. The generic profile showing DA^2 admits a NE leading to an unstable outcome is given in Claim 8.

Claim 8. *For any k, there exists a profile (given in Example 14) such that a NE outcome of DA^k is unstable.*

Proof. The generic example is the following:

Example 14.

P_1 :	s_2	$\boxed{s_1}$	( )
P_2 :	$\boxed{s_3}$	s_2	( )
P_3 :	$\boxed{s_2}$	s_3	( )
$\vdots$			

F_{s_1} :	t_1	( )		
F_{s_2} :	t_2	t_1	t_3	( )
F_{s_3} :	t_3	t_2	( )	
$\vdots$				

▲

Any profile in DA^k such that

- $Q_1 : s_1$ ( ), $Q_2 : s_3\ s_2$ ( ) and $Q_3 : s_2\ s_3$ ( ),
- $Q_{-1,2,3}$ is a NE in DA^k for the sub-profile M' with
$$M' := (F\backslash\{F_{s_1}, F_{s_2}, F_{s_3}\}, q\backslash\{q_{s_1}, q_{s_2}, q_{s_3}\}, P\backslash\{P_1, P_2, P_3\})$$

is a NE and leads to an outcome $DA^k(Q)$ (boxed) for which the pair (t_1, s_2) is blocking. Such a NE in the sub-profile always exists and the profile $Q := (Q_1, Q_2, Q_3, Q_{-1,2,3})$ is a NE in the profile (F, q, P). Indeed, t_2 and t_3 are assigned to their favorite school, t_1 is assigned to her second favorite school and there exists a rejection chain preventing t_1 to be assigned to s_2, if she declared s_2. Furthermore, no $t_i \in T\backslash\{t_1, t_2, t_3\}$ can obtain a school in $\{s_1, s_2, s_3\}$ given Q_1, Q_2 and Q_3 and F.

■

Proof of Proposition 15
(DA^k more stable than BOS^k in US)

Lemma 20. *A strategy Q_i is safe for t_i in BOS^k, if and only if*

(i) $Q_i(1)$ is safe if favorite, or

(ii) there exists a set of over-supplied schools $O \subset Q_i$ (i.e. $\sum_{s\in O} q_j \geq n$) with $|O| \leq k$.

Proof. Sufficiency is obvious, so we focus on necessity. We prove the contrapositive. Assume neither (i) nor (ii) are true. Consider any sub-profile Q^*_{-i} constructed as follows

- Take any set of $q_{Q_i(1)}$ students $t_j \neq t_i$ among the students with higher priority at $Q_i(1)$ than t_i, and let
$$Q^*_j(1) := Q_i(1).$$
- For any $\ell \in \{2, \ldots, |Q_i|\}$ take $q_{Q_i(\ell)}$ students t_j whose declaration has not been constrained yet and let
$$Q^*_j(1) := Q_i(\ell).$$

Because (i) is false, there are at least $q_{Q_i(1)}$ students in T with higher priority at school $Q_i(1)$ than t_i.

Because (ii) is false, any oversupplied set of schools contains more than k schools. Therefore $\{Q_i(1), \ldots, Q_i(|Q_i|)\}$ is not an oversupplied set of schools.

There are hence enough students to perform the constructions described above. Therefore Q^*_{-i} is well-defined. Clearly, by construction and (i) we have $BOS^k_i(Q_i, Q^*_{-i}) \neq Q_i(1)$. By construction again, for every $s \neq Q_i(1) \in Q_i$, there are at least q_s students who apply to s in the first round of BOS^k. Therefore $BOS^k_i(Q_i, Q^*_{-i}) \neq s$ and we have $BOS^k_i(Q_i, Q^*_{-i}) = t_i$, showing Q_i is not a safe strategy.

■

Lemma 21. *If Q_i and Q'_i are different unsafe strategy in BOS^k and Q'_i contains $\min\{k, |P_i|\}$ schools, all schools in Q'_i being acceptable, then Q_i does not dominate Q'_i.*

Proof. By contradiction. Let r be the lowest rank for which $Q_i(r) \neq Q'_i(r)$.

If $r > |Q'_i|$, implying that strategy Q'_i is the truncation of Q_i after rank $|Q'_i|$, then we have $|Q'_i| = |P_i| < k$. As a consequence, $Q_i(r)$ is unacceptable because all acceptable schools are declared in Q_i before rank r. By Lemma 16, there exists Q^*_{-i} such that $BOS^k_i(Q_i, Q^*_{-i}) = Q_i(r)$, that is t_i is assigned an unacceptable school. As all schools declared in Q'_i are acceptable, t_i strictly prefers her assignement when declaring Q'_i and other students declare Q^*_{-i}. This shows that strategy Q_i does not dominate Q'_i.

There remains to consider cases for which $r \leq |Q'_i|$. To obtain a contradiction, we construct Q^*_{-i} such that

$$BOS^k_i(Q'_i, Q^*_{-i}) \ P_i \ BOS^k_i(Q_i, Q^*_{-i}) = t_i,$$

that is t_i is assigned to an acceptable school when playing Q'_i and unassigned for Q_i. We consider two cases for the construction.

First, if $Q'_i(r) \in Q_i$:

- Take the $q_{Q_i(1)}$ students $j \neq i$ with highest priority at school $Q_i(1)$ an let $Q^*_j : Q_i(1)$,
- For all $s \in Q_i$ with $s \neq Q_i(1)$ and $s \neq Q'_i(r)$, take q_s students u whose declaration is not yet constrained and let $Q^*_u : s$,
- Take $q_{Q'_i(r)} - 1$ students v whose declaration is not yet constrained and let $Q^*_v : Q'_i(r)$,
- Take a student g not yet constrained. If $t_i \ F_{Q'_i(r)} \ t_g$ then Q^*_g is the truncation of Q'_i after school $Q'_i(r)$, else it is the truncation of Q_i after school $Q_i(r)$ with in addition $Q_g(r+1)^* := Q'_i(r)$.
- Students whose preference is not specified do not declare school $Q'_i(r)$.

It is possible to construct this Q^*_{-i}. First, Q_i is unsafe and has hence no SIF in $Q_i(1)$. As a result, there are enough students j. Second, there are enough students to construct Q^*_{-i} because the number of students whose preference is constrained (including student t_i) is equal to the sum of seats available in schools declared in the unsafe Q_i. As there is no over-supplied set of schools, Q^*_i can be constructed.

By construction, $BOS^k_i(Q_i, Q^*_{-i}) = t_i$ as all seats in all schools declared in Q_i are allocated at step 1 of the algorithm to other students than t_i, except one in school $Q'_i(r)$ if $r \neq 1$. This last seat is allocated to t_i at step r for strategy Q'_i and is allocated to student g at step r or $r+1$ for strategy Q_i.

The second case is $Q'_i(r) \notin Q_i$. The construction of Q^*_{-i} is almost identical. The only difference is that no student v is constrained to declare $Q^*_v : Q'_i(r)$ and student g's preference is not constrained. ■

Proposition 17 (US outcomes of DA^k nested in US outcomes of BOS^k)**.** *Take any Q, a US of DA^k. Assume that given k, no set of oversupplied schools can be declared in DA^k or BOS^k, that is*

$$\textit{there exists no } O \subset S \textit{ with } |O| \leq k \textit{ such that } \sum_{s \in O} q_s \geq |T|. \tag{3.46}$$

Then, there exists Q'' a US of BOS^k such that $DA^k(Q) = BOS^k(Q'')$.

Proof: First notice that by Lemma 13, Lemma 14 and Lemma 15, it is enough to prove that $BOS^k(Q'') = DA^k(Q')$, where for all $t_i \in T$, Q'_i ranks exactly $\min\{k, |P_i|\}$ acceptable schools without re-ranking, and no other schools are ranked (in particular no unacceptable schools are ranked).

The proof consists in showing that any profile Q'' in which

(i) For all t_i who is *un*assigned in $DA^k(Q')$,
$$Q''_i := Q'_i.$$

(ii) For all t_i who is assigned in $DA^k(Q')$,
$$Q''_i(1) := DA^k_i(Q'),$$
and Q''_i ranks $\min\{k, |P_i|\}$ acceptable schools, *including all the (acceptable) schools in Q'_i.*[21]

is an undominated strategy of BOS^k. Once this is proven, it is easy to see that $BOS^k(Q'') = DA^k(Q')$, which yields the desired result desired result.

(i) yields an US for all t_i *un*assigned in $DA^k(Q')$. We first show that Q''_i is unsafe. Consider the two cases in Lemma 20. By assumption, case (ii) is ruled out. Therefore, Q''_i is safe if and only if $Q''_i(1)$ is SIF. But because $Q''_i(1) = Q'_i(1)$, if $Q''_i(1)$ was SIF, we would have $DA^k(Q) = Q_i(1)$, contradicting the assumption that t_i is unassigned in $DA^k(Q)$. Therefore, Q''_i is unsafe.

Now, in order to derive a contradiction, assume that Q'''_i dominates Q''_i in BOS^k. By construction, $Q''_i = Q'_i$ ranks $\min\{k, |P_i|\}$ acceptable schools. Hence, by Lemma 21, Q''_i cannot be dominated by an unsafe strategy like Q'''_i.

We obtain a contradiction by showing that Q'''_i cannot be safe either. Again by Lemma 20 and the assumption on oversupplied schools, if Q'''_i is safe, $Q'''_i(1)$ is SIF. This in turn means $Q'''_i(1) \notin Q''_i$. Indeed, because $Q'''_i(1)$ is SIF and $Q''_i = Q'_i$, t_i could not be unassigned in $DA^k(Q')$. Because Q''_i is unsafe (see above), Lemma 16 applies and for all $\ell \in \{1, \ldots, |Q''_i|\}$, there exists Q^ℓ_{-i} such that
$$BOS^k_i(Q''_i, Q^\ell_{-i}) = Q''_i(\ell).$$
But because $Q'''_i(1)$ is a SIF and Q'''_i dominates Q''_i in BOS^k, the last displayed equality implies
$$Q'''_i(1) \; R_i \; Q''_i(\ell), \qquad \text{for all } \ell \in \{1, \ldots, |Q''_i\}.$$

[21] This is feasible since $DA^k_i(Q')$ is acceptable by construction of Q'_i.

Finally, because $Q_i'''(1)$ is not declared in Q_i'', the last displayed relation implies

$$Q_i'''(1)\ P_i\ Q_i''(\ell), \qquad \text{for all } \ell \in \{1, \dots, |Q_i''\},$$

and because $Q_i'' = Q_i'$, this means Q_i''' would also dominate Q_i' in DA^k, contradicting the fact that Q_i' is an US of DA^k. Hence, Q_i''' is unsafe, a contradiction to Lemma 21.

(ii) yields an US for all t_i assigned in $DA^k(Q')$.

Case 1 : $DA_i^k(Q')$ is SIF.

By construction, $DA_i^k(Q')$ is acceptable. Therefore, Q_i'' can only be dominated by a safe strategy. But by Lemma 20 and the assumption on oversupplied schools, any safe strategy Q_i''' that would dominate Q_i'' would be such that $Q_i'''(1)$ is SIF. Thus, this would mean that there exists a SIF $Q_i'''(1)$ such that

$$Q_i'''(1)\ P_i\ Q_i''(1).$$

But because $Q_i''(1) = DA_i^k(Q')$ and Q' is without re-ranking by construction, this implies $Q_i'''(1) \notin Q_i'$ as otherwise, we would have $DA_i^k(Q') = Q_i'''(1) \neq Q_i''(1)$, a contradiction. This in turn implies that Q_i' is dominated in DA^k by a strategy Q_i^* constructed from Q_i' by only replacing $Q_i''(1)$ by $Q_i'''(1)$, contradicting the assumption that Q_i' is an US of DA^k.

Case 2 : $DA_i^k(Q')$ is not SIF.

Again, by Lemma 20 and the assumption on oversupplied schools, Q_i'' is unsafe. But because Q_i'' ranks $\min\{k, |P_i|\}$ acceptable schools, Lemma 21 applies and any Q_i''' dominating Q_i'' must be a safe strategy. Now, by the same argument as in (i), this implies $Q_i'''(1)$ is a SIF that t_i strictly prefers to all the schools in Q_i', contradicting the assumption that Q_i' is an US of DA^k.

We showed that US outcomes in DA^k are also US outcomes in BOS^k. As a result, if all US outcomes in BOS^k are stable, then all US outcomes in DA^k are stable as well. This implies that DA^k is at least as stable as BOS^k. Next, we show that DA^k is more stable than BOS^k. Examples 8 and 9 in the text provide each a profile for the comparison of DA^2 and BOS^2. The generic profile required is given in Claim 9.

Claim 9. *For any k, there exists a profile (given in Example 15) such that all US outcomes are stable in DA^k whereas some US outcomes are unstable in BOS^k.*

Proof. The generic example is the following:

Example 15.

$$\begin{array}{llll} P_1: & s_1^* & \dots & s_k \\ P_2: & s_1 & \dots & s_k \\ \vdots & & & \\ P_k: & s_1 & \dots & s_k^* \\ P_{k+1}: & s_k & & \end{array} \qquad \begin{array}{lllll} F_{s_1}: & t_1 & \dots & t_k & t_{k+1} \\ F_{s_2}: & t_1 & \dots & t_k & t_{k+1} \\ \vdots & & & & \\ F_{s_k}: & t_1 & \dots & t_k & t_{k+1} \end{array}$$

▲

All students except t_{k+1} have the same preferences. All schools have the same priority rankings. The unique stable outcome is such that for all $i \in \{1, \ldots, k\}$, student t_i is assigned to s_i and t_{k+1} is unassigned.

There are many reasons why BOS^k is unstable in US. One of them is that there exists Q_k undominated in BOS^k such that $Q_k(1) \neq s_k$. If this US is declared, because $Q_{k+1} : s_k$ is the only undominated strategy of t_{k+1}, t_k ends up unassigned, although she has higher priority at s_k than t_{k+1}.

On the other hand, DA^k is stable in US since any t_i with $i \in \{1, \ldots, k\}$ must declare, in any undominated strategy, her i preferred schools first without re-rankings. This is because t_i has a safe set covering her i preferred schools.

■

■

3.9.5 Other proofs and examples

Results about dominant strategies

Proposition 18 (Characterization of dominant strategies in DA^k)**.** *Q_i is a dominant strategy in DA^k if and only if either*

(i) $Q_i = P_i$, or

(ii) the $q \leq k$ preferred schools in P_i form a safe set that is declared without re-ranking in Q_i.

Proof. Sufficiency.

(i) This follows directly from the strategy-proofness of DA^m. Because DA^m is strategy-proof, $Q_i := P_i$ is a dominant strategy in DA^m, that is

$$DA_i^m(P_i, Q_{-i}) \; P_i \; DA_i^m(Q_i', Q_{-i}), \qquad \text{for all } Q_i' \text{ and all } Q_{-i}.$$

In particular,

$$DA_i^m(P_i, Q_{-i}^k) \; P_i \; DA_i^m(Q_i^{k\prime}, Q_{-i}^k), \text{ for all } Q_i^{k\prime} \text{ with } |Q_i^{k\prime}| \leq k$$
$$\text{and all } Q_{-i}^k \text{ with } |Q_j^k| \leq k \text{ for all } t_j \neq t_i.$$

But because DA^m is equivalent to DA^k when students declare no more than k schools, the last displayed relation is equivalent to

$$DA_i^k(P_i, Q_{-i}^k) \; P_i \; DA_i^k(Q_i^{k\prime}, Q_{-i}^k), \text{ for all } Q_i^{k\prime} \text{ with } |Q_i^{k\prime}| \leq k$$
$$\text{and all } Q_{-i}^k \text{ with } |Q_j^k| \leq k \text{ for all } t_j \neq t_i,$$

hence P_i is a dominant strategy in DA^k.

(ii) In order to derive a contradiction, assume there exists some Q_i^* and Q_{-i}^* such that

$$DA_i^k(Q_i^*, Q_{-i}^*) \; P_i \; DA_i^k(Q_i, Q_{-i}^*). \tag{3.47}$$

Following the same argument as in point (i), Q_i would be a dominant strategy if the true preferences were Q_i, that is

$$DA_i^k(Q_i, Q_{-i})\ Q_i\ DA_i^k(Q'_i, Q_{-i}), \qquad \text{for all } Q'_i \text{ and all } Q_{-i}.$$

Because Q_i is without re-ranking and ranks all schools up to $P_i(q)$, the last displayed relation implies

$$DA_i^k(Q_i, Q_{-i})\ R_i\ DA_i^k(Q'_i, Q_{-i}), \qquad \text{for all } Q'_i \text{ and all } Q_{-i} \text{ such that } DA^k(Q'_i, Q_{-i})\ R_i\ P_i(q). \tag{3.48}$$

Now by definition of a safe set,

$$DA_i^k(Q_i, Q_{-i})\ Q_i\ P_i(q),$$

and again, because Q_i is without re-ranking

$$DA_i^k(Q_i, Q_{-i})\ P_i\ P_i(q).$$

Therefore,

$$DA_i^k(Q_i, Q_{-i})\ R_i\ DA_i^k(Q'_i, Q_{-i}), \qquad \text{for all } Q'_i \text{ and all } Q_{-i} \text{ such that } P_i(q)\ P_i\ DA_i^k(Q'_i, Q_{-i}). \tag{3.49}$$

Together (3.48) and (3.49) imply

$$DA_i^k(Q_i, Q_{-i})\ R_i\ DA_i^k(Q'_i, Q_{-i}), \qquad \text{for all } Q'_i \text{ and all } Q_{-i}$$

In particular,

$$DA_i^k(Q_i, Q^*_{-i})\ R_i\ DA_i^k(Q^*_i, Q^*_{-i}).$$

contradicting (3.47).

Necessity. By contradiction. Assume there exists a dominant strategy Q_i that violates both (i) and (ii). We consider two alternative cases.

Case 1 : Q_i is an unsafe strategy.

We first show that in this case, *there exists an acceptable school s^* which is not declared in Q_i*. The proof of the existence of s^* is based on proofs showing that a dominant unsafe strategy contains no unacceptable schools and no re-rankings.

Because Q_i is a dominant strategy of DA^k, it is also a US of DA^k. Thus, by Lemma 13 and Lemma 14, we can assume without loss of generality that Q_i is without re-ranking and only contains acceptable school. Thus, the only way to violate (i) and have $Q_i \neq P_i$ is if there exists an acceptable school s^* not declared in Q_i.

As Q_i is unsafe, there exists Q_{-i} such that $DA_i^k(Q_i, Q_{-i}) = t_i$. As $s^* \notin Q_i$, we have $t_i\ Q_i\ s^*$. Therefore Lemma 12 applies: there exists a profile Q^*_{-i} such that $DA_i^k(Q_i, Q^*_{-i}) = t_i$ and Q^*_i such that $DA_i^k(Q^*_i, Q^*_{-i}) = s^*$. Together we have

$$DA_i^k(Q^*_i, Q^*_{-i})\ P_i\ DA_i^k(Q_i, Q^*_{-i}),$$

contradicting the assumption that Q_i is a dominant strategy.

Case 2 : Q_i is a safe strategy.

First, notice that in this case, there exists a school s^* with rank $r^* \leq k$ in P_i such that:

1. s^* is not declared in Q_i,
2. the least preferred school declared in $PAS(Q_i)$, denoted $w^{PAS(Q_i)}$ is such that:

$$s^* \; P_i \; w^{PAS(Q_i)}.$$

Indeed, by Lemma 13 and Lemma 14, we can again assume without loss of generality that Q_i is without re-ranking and only contains acceptable school. Then, the only way for (ii) to be violated is if there exists a school s^* with the properties defined above.

As $w^{PAS(Q_i)} \in PAS(Q_i)$, there exists Q_{-i} such that $DA_i^k(Q_i, Q_{-i}) = w^{PAS(Q_i)}$. As $s^* \notin Q_i$, we have also that $w^{PAS(Q_i)} \; Q_i \; s^*$. Therefore Lemma 12 applies: there exist Q^*_{-i} such that $DA_i^k(Q_i, Q^*_{-i}) = w^{PAS(Q_i)}$ and Q^*_i such that $DA_i^k(Q^*_i, Q^*_{-i}) = s^*$. Together we have

$$DA_i^k(Q^*_i, Q^*_{-i}) \; P_i \; DA_i^k(Q_i, Q^*_{-i}),$$

contradicting the assumption that Q_i is a dominant strategy.

■

Proposition 19 (Dominant strategy implies not part of blocking pair in DA^k)**.** *For all Q_{-i}, if Q_i is dominant strategy in DA^k, then t_i does not participate to a blocking pair in $DA^k(Q_i, Q_{-i})$.*

Proof. By contradiction. Assume that for some profile Q_{-i}, student t_i is blocking at a school s in allocation $DA^k(Q_i, Q_{-i})$. By the definition of a blocking pair, this implies that $s \; P_i \; DA_i^k(Q_i, Q_{-i})$ and there exists t_g with $DA_g^k(Q) = s$ and $t_i \; F_s \; t_g$. By Proposition 18, two alternative cases must be considered:

- Case 1: $Q_i = P_i$.

  As s is declared in Q_i before school $DA_i^k(Q_i, Q_{-i})$, since $s \; P_i \; DA_i^k(Q_i, Q_{-i})$, student t_i was rejected from s in the proceeding of mechanism DA^k. At the step at which t_i is rejected from s, there are q_s students $t_j \neq t_i$ assigned to s with higher priority at s than t_i. If any student t_j is rejected from s in a later step of DA^k, the seat in s previously occupied by t_j is assigned to another student t_k with higher priority at s than t_j, and hence with higher priority at s than t_i. There are hence no student t_g with $t_i \; F_s \; t_g$ such that $DA_g^k(Q) = s$, contradicting the assumption that t_i was blocking at s.

- Case 2: the $q \leq k$ preferred schools in P_i form a safe set that is declared without re-ranking in Q_i.

  By the definition of a safe set, this case implies that for all Q_{-i} we have

$$DA_i^k(Q_i, Q_{-i}) \; P_i \; P_i(q+1).$$

As t_i is blocking at school s, we have $s\ P_i\ DA_i^k(Q_i, Q_{-i})$. As we showed $DA_i^k(Q_i, Q_{-i})\ P_i\ P_i(q+1)$, this case is such that s is declared in Q_i before $DA_i^k(Q_i, Q_{-i})$. The argument is then the same as in Case 1.

■

Proposition 20 (Dominant strategy implies not part of blocking pair in BOS^k)**.** *Assume there is no set of over-supplied schools O with $|O| \leq k$. Strategy Q_i is a dominant strategy in BOS^k with $k \geq 3$ if and only if $Q_i(1) = P_i(1)$ and either*

1. *$P_i(1)$ is SIF, or*
2. *$|P_i| = |Q_i| = 1$.*

Proof. The *sufficiency* of these conditions is a direct corollary of Proposition 18 (this proposition is applicable as we assume there is no set of over-supplied schools). When $|P_i| = 1$ and $P_i(1)$ is not SIF, the *necessity* of condition 2 is a corallary of the fact that a single strategy ($Q_i := P_i$) qualifies to be US in Proposition 18. When $|P_i| \geq 2$, we show the *necessity* of condition 1 for strategy Q'_i to be a dominant strategy. Two cases must be considered

- Case 1: $P_i(1)$ is SIF but $Q'_i(1) \neq P_i(1)$.

  By proposition 18, such Q'_i is not an US in BOS^k and hence not a dominant strategy.

- Case 2: $P_i(1)$ is not SIF.

  We prove for this case the non-existence of dominant strategy by showing the existence of Q'_i and Q''_i such that

  - Q'_i and Q''_i are US in BOS^k and
  - Q'_i and Q''_i are not equivalent strategies.

  Let Q'_i be such that $Q'_i(1) := P_i(1)$ and Q'_i contains $min\{k, |P_i|\}$ acceptable schools. Strategy Q'_i is US by Proposition 18. Let Q''_i be such that $Q''_i(1) := P_i(2)$ and Q''_i contains $min\{k, |P_i|\}$ acceptable schools. Strategy Q''_i is US by Proposition 18 because $|P_i| \geq 2$ and $P_i(1)$ is not SIF. Observe this is true whether $P_i(2)$ is SIF or not.

  There exists Q^*_{-i} – for example Q^*_j contains no school for all $t_j \neq t_i$ – for which the two US Q'_i and Q''_i yield different assignements and they are hence not equivalent.

■

Proposition 21. *Assume there is no set of over-supplied schools O with $|O| \leq k$. For all $k \geq 3$ and all Q_{-i}, if Q_i is a dominant strategy in BOS^k then t_i does not participate to a blocking pair in $BOS^k(Q)$.*

Proof. By contradiction. Assume t_i participates to a blocking pair with school s. This implies by definition that $s \; P_i \; BOS_i^k(Q)$. By Lemma 20, if Q_i is a dominant strategy then two cases can arise:

- Case 1: $P_i(1)$ is SIF and $Q_i(1) = P_i(1)$.
  This implies $BOS_i^k(Q) = P_i(1)$. There exists hence no s with $s \; P_i \; BOS_i^k(Q)$, contradicting our assumption.

- Case 2: $|P_i| = 1$ and $Q_i = P_i$.
  If $BOS_i^k(Q) = P_i(1)$, then t_i can not participate to a blocking pair for the reason explained in Case 1. If on the other hand $BOS_i^k(Q) = t_i$, then student t_i was rejected from $P_i(1)$ in the first step of $BOS^k(Q)$, implying that $q_{P_i(1)}$ students with higher priority at $P_i(1)$ than t_i are assigned to $P_i(1)$. Therefore t_i can not participate to a blocking pair as she only finds $P_i(1)$ acceptable.

■

When Alternative criterion 4 does not apply

We start with a short reminder of notation that will be useful for this example. For assignment profile M, mechanism ϕ and solution concept C, let $C^\phi(M)$ be the set of C-profiles under ϕ. We denote the cardinality of such set using a lower-case $c^\phi(M)$. Let $C^\phi_{Stab}(M)$ be the subset of C-profiles under ϕ whose associated outcome satisfies stability, and $C^\phi_{\neg Stab}(M)$ be the subset of C-profiles which do not. Similarly, the lower case $c^\phi_{Stab}(M)$ and $c^\phi_{\neg Stab}(M)$ denote the cardinality of the corresponding sets.

Our criterion concludes that DA^2 is more stable than BOS^2 in US (Proposition 15). We show using Example 16 that Alternative criterion 4 is silent when comparing these two mechanisms.

We show the existence of two type profiles M_1 and M_2 exhibiting no over-supplied set of schools such that:

$$\frac{us^{BOS^2}_{Stab}(M_1)}{us^{BOS^2}_{Stab}(M_1) + us^{BOS^2}_{\neg Stab}(M_1)} > \frac{us^{DA^2}_{Stab}(M_1)}{us^{DA^2}_{Stab}(M_1) + us^{DA^2}_{\neg Stab}(M_1)}, \tag{3.50}$$

$$\frac{us^{BOS^2}_{Stab}(M_2)}{us^{BOS^2}_{Stab}(M_2) + us^{BOS^2}_{\neg Stab}(M_2)} < \frac{us^{DA^2}_{Stab}(M_2)}{us^{DA^2}_{Stab}(M_2) + us^{DA^2}_{\neg Stab}(M_2)}. \tag{3.51}$$

The existence of these two type profiles pointing towards different comparisons renders Alternative criterion 4 silent.

The existence of M_2 is shown in Example 8 for which all US profiles in DA^2 lead to stable outcomes whereas some US profiles in BOS^2 lead to unstable outcomes. The existence of M_1 is shown by the profile given in Example 16.

Example 16.

$$\begin{array}{lllll} P_1: & s_2 & s_1^* & & \\ P_2: & s_1 & s_2^* & & \\ P_3: & s_3^* & (\,) & & \\ P_4: & s_4^* & (\,) & & \\ P_5: & s_5^* & (\,) & & \\ P_6: & s_1 & s_3 & s_4 & s_5 \end{array} \qquad \begin{array}{lllll} F_{s_1}: & t_1 & t_6 & t_2 & (\,) \\ F_{s_2}: & t_2 & t_1 & (\,) & \\ F_{s_3}: & t_3 & (\,) & & \\ F_{s_4}: & t_4 & (\,) & & \\ F_{s_5}: & t_5 & (\,) & & \end{array}$$

▲

Observe that for $k = 2$, there is no set of over-supplied school in M_1. The most efficient stable assignment is starred. Observe in M_1 that if t_2 is assigned to school s_1 then student t_6 is blocking at s_1. The proof of inequality (3.50) is in two steps.

Step 1 :

$$\frac{us^{DA^2}_{Stab}(M_1)}{us^{DA^2}_{Stab}(M_1) + us^{DA^2}_{\neg Stab}(M_1)} = \frac{1}{2}.$$

The unique undominated strategy in DA^2 for students t_1 and t_2 is to reveal truthfully $Q_1 := P_1$ and $Q_2 := P_2$ as both t_1 and t_2 have a safe set covering their two preferred schools. The dominant strategy for t_3, t_4 and t_5 is to declare their favorite school first as each of these students has highest priority at her favorite school. For student t_6, any strategy Q_6 revealing two of her four acceptable schools without re-rankings is US. There are 6 such strategies:

$$\begin{array}{lll} Q_6^1: & s_1 & s_3 \\ Q_6^2: & s_1 & s_4 \\ Q_6^3: & s_1 & s_5 \end{array} \qquad \begin{array}{lll} Q_6^4: & s_3 & s_4 \\ Q_6^5: & s_3 & s_5 \\ Q_6^6: & s_4 & s_5 \end{array}$$

and hence 6 US profiles $(Q_1, Q_2, Q_3, Q_4, Q_5, Q_6^x)$ in DA^2. One can easily see that strategies Q_6^4, Q_6^5 and Q_6^6 lead to unstable outcomes as t_6 does not declare s_1 and consequently t_2 is assigned to s_1, which makes t_6 blocking at s_1. The three other undominated strategies of t_6 lead to stable outcomes, the desired result.

Step 2 :

$$\frac{us^{BOS^2}_{Stab}(M_1)}{us^{BOS^2}_{Stab}(M_1) + us^{BOS^2}_{\neg Stab}(M_1)} > \frac{1}{2}.$$

For student t_1, any undominated strategy Q_1 in BOS^2 is such that either $Q_1(1) = s_2$, or $Q_1(1) \neq s_2$. When $Q_1(1) = s_2$, by Lemma 18, the only US is the truthful revelation $Q_1 :\ s_2\ s_1$. On the other hand, when $Q_1(1) \neq s_2$ in a US, $Q_1(1) = s_1$ and the two possible US are $Q_1 : s_1$ and $Q_1 : s_1\ s_2$. As a result, student t_1 declares school s_1 to be her favorite in at least half of the US profiles in BOS^2.[22]

[22] This holds even if $Q_1 : s_1$ and $Q_1 : s_1\ s_2$ are viewed as equivalent strategies.

We show that

1. any US profile for which t_1 declare school s_1 as her favorite school lead to a stable outcome.
2. there exists US profile Q^* for which $Q_1^* :\ s_2\ s_1$ and Q^* leads to a stable outcome.

As student t_1 declare school s_1 to be her favorite in at least half of the US profiles in BOS^2, (1) and (2) imply the desired result. We prove these claims in turn.

1. By Lemma 18, students t_3, t_4 and t_5 declare their favorite school first in any undominated strategy of BOS^2 and are assigned to this school because their favorite school is SIF.

   By Lemma 18, any undominated strategies Q_2 of t_2 is such that either

   - $Q_2' = P_2$, or
   - $Q_2'(1) = s_2$.

   In both cases, if student t_1 declares s_1 to be her favorite, student t_2 is assigned to s_2. Indeed, except for t_2, only student t_1 declares s_2.

   Finally, the declaration of student t_6 does not influence the outcome as all acceptable schools of t_6 are assigned in the first step of BOS^2 to a student declaring this school first (this school is SIF for the student declaring it first).

   This shows that when student t_1 declare s_1 to be her favorite, all US profile lead to the most efficient stable assignment.

2. The following profile Q^* is such that $Q_1^* :\ s_2\ s_1$. Furthermore, Q^* is US in BOS^k by Lemma 18 and leads to the most efficient stable assignment.

$Q_1^* :$	$s_2 \quad s_1$	$Q_4^* :$	s_4
$Q_2^* :$	s_2	$Q_5^* :$	s_5
$Q_3^* :$	s_3	$Q_6^* :$	$s_3 \quad s_4$

Chapter 4

Disambiguation of Ellsberg equilibria in 2×2 normal form games

(Joint with Frank Riedel)

4.1 Introduction

The presence of ambiguity in strategic interactions has recently received increasing attention. There have been attempts to allow for ambiguous acts or beliefs in games. For complete information normal form games, two strands of literature can be distinguished. The first strand, which has been introduced by Lo (1996), Marinacci (2000) and Eichberger and Kelsey (2000), considers *subjective* ambiguity. Ambiguity is introduced in the beliefs players hold about the strategies adopted by their opponents. This strand extends the belief interpretation of Nash equilibria by allowing for equilibria in ambiguous beliefs. A disadvantage of these equilibria is that they leave unanswered the question of which strategy profile is played in equilibrium. The second strand, introduced more recently by Riedel and Sass (2013), considers *objective* ambiguity. The set of available strategies is expanded to ambiguous randomization strategies – called Ellsberg strategies – which are convex sets in the space of mixed strategies. Players may therefore render their strategy objectively ambiguous. Riedel and Sass (2013) call such an extended game an Ellsberg game. The solution concept proposed is the Ellsberg equilibrium: players play a best response to the Ellsberg strategy of their opponent. As a consequence, the Ellsberg equilibrium is a more general solution concept than the Nash equilibrium.

In Ellsberg games, existence of Ellsberg equilibria follows from the existence of Nash equilibria. Riedel and Sass (2013) have shown that in addition to the Nash equilibria, new Ellsberg equilibria may arise in which players use proper Ellsberg strategies. Interestingly, in games with at least three players, some of these new equilibria yield outcomes that cannot be reached under Nash equilibria. In other words, their solution concept expands the support of the outcomes. Riedel and Sass (2013) show this last point by means of an example taken from Greenberg (2000). In the example, two small countries decide for themselves whether to engage in a war against each other, or to stay at peace. A superpower tries to negotiate for peace by threatening to punish one of them in case war breaks out. Being unable to identify which country is responsible when war breaks out, the superpower's best reply is to punish one of the two countries picked at random with probability one half. As a result, the only Nash equilibrium has the small countries engage in war. If the superpower had the possibility to "remain silent" and could be sufficiently ambiguous about which country it would punish, a new Ellsberg equilibrium would appear, with peace as the unique outcome. Because the two small countries are pessimistic in the face of ambiguity, each country assigns a high probability to being punished in case of war, and hence do not engage in a war. Greenberg argues that such outcome would be more realistic.

We provide an alternative interpretation for Ellsberg equilibria. Ellsberg equilibria generalize mixed strategy equilibria. Mixed strategies play a central role in game theory. Without mixing, it would for instance be impossible to assign values to zero-sum games or to find Nash equilibria in more general strategic interactions.

The classic interpretation of a mixed strategy was introduced by John von Neumann and relies on the use of an objective randomization device. Riedel and Sass (2013) put forward a direct generalization of von Neumann's idea to ambiguous strategies by allowing players to use Ellsberg urns with given parameters. That is, players base their actions on the outcome of an Ellsberg urn experiment where the probabilities are only known up to some bounds. While such a construction makes perfect sense in theory, one might wonder whether it would be implementable in actual games.

Even in its classic form (i.e. not allowing for ambiguity and the use of Ellsberg urns), the objective randomization device interpretation has been questioned and criticized. While deliberate use of a random device makes sense in a strictly competitive game (Neumann, 1928), it might be more questionable in more cooperative situations like a coordination game (Schelling, 1980).

There exists however an interesting alternative interpretation of mixed strategies. Harsanyi (1973) has shown that mixed strategy equilibria may be viewed as limits of pure strategy equilibria in a slightly disturbed game where players have private information about their payoffs. In this paper, we show how one can purify, or at least disambiguate, Ellsberg equilibria in the spirit of Harsanyi's approach. We show that Ellsberg equilibria can be viewed as limits of equilibria in slightly disturbed games where the disturbances are ambiguous. The Disambiguation Theorem we prove is an extension of Harsanyi's Purification Theorem.

We confine our analysis to two-players games with two actions for each player. We identify one class of games where one can purify the Ellsberg equilibria. As in Harsanyi (1973), the players use pure strategies of a threshold type in the disturbed version of the game. From the perspective of an outside observer, these actions induce, in the limit, the same set of probability distributions as the Ellsberg equilibrium.

For the games outside this class, we *disambiguate* the Ellsberg equilibria in the following way. In the disturbed games, players best reply using their two pure strategies *and their maxmin strategy*. The maxmin strategy plays a key role in Ellsberg games as it allows players to hedge against Knightian uncertainty. The appearance of such maxmin strategies is therefore natural in our context. Again, we can show that from the perspective of an outside observer, the induced distributions of actions coincide in the limit with the distributions of the Ellsberg equilibrium.

We draw the reader's attention to the limitations of the Disambiguation Theorem we prove. It is less general than Harsanyi's theorem in two ways. First, its scope is limited to 2×2 normal form games, whereas Harsanyi's theorem is valid for all finite n-player non-cooperative games. Harsanyi's technique cannot be adapted to the case of multiple priors we study because it relies on smoothness of the payoff functions, which is lost when one uses a multiple prior representation for preferences. Second, in our setting, the payoffs associated to a given strategy are subject to identical disturbances, whereas disturbances are independent in Harsanyi's setting.

The paper is organized in three parts. In section 4.2, we introduce the definitions and notation. In section 4.3, we present and prove our Disambiguation Theorem. Finally, in section 4.4, we provide an example of disambiguation for a particular 2×2 coordination game.

4.2 Definitions and notation

We first present the basic 2×2 normal form Ellsberg game. We first describe the ambiguous randomization strategies available to players, how these players behave in the face of ambiguity and provide the definition of Ellsberg equilibria. Then, we delimit the class of games considered. Finally, we describe the disturbed versions of the basic game, the strategies available to players in these disturbed versions, and we show that these strategies are perceived as Ellsberg strategies by external observers.

4.2.1 The basic game Γ

The games we consider are 2×2 normal form games, illustrated in Figure 4.1. Basic notation and definitions for these games are as follows:

- Let $p, q \in [0, 1]$ denote the mixed strategy of player 1 and 2 respectively.
- Let the pair $(p, q) \in [0, 1] \times [0, 1]$ denote a mixed strategy profile.
- Player i's expected utility for the strategy profile (p, q) is $U_i(p, q)$ with
$$U_i(p, q) = pq\pi_i^1 + p(1-q)\pi_i^2 + (1-p)q\pi_i^3 + (1-p)(1-q)\pi_i^4.$$
- For player 1, strategy p is a best reply to q if $U_1(p, q) \geq U_1(p', q)$ for all $p' \in [0, 1]$.
- Strategy profile (p, q) is a Nash equilibrium if p and q are mutual best replies.

		Player 2: L (q)	Player 2: R (1-q)
Player 1	U (p)	$\pi_1^1; \pi_2^1$	$\pi_1^2; \pi_2^2$
	D (1-p)	$\pi_1^3; \pi_2^3$	$\pi_1^4; \pi_2^4$

Figure 4.1: Normal form of the basic 2×2 game Γ.

In Ellsberg games, in addition to pure and mixed strategies, players can use ambiguous randomization strategies called *Ellsberg strategies*. For 2×2 games, an Ellsberg strategy is a closed interval $[a, b]$ in the probability space $[0, 1]$. If player 1 plays an Ellsberg strategy, she plays the pure strategy U with a probability inside $[a, b]$, but the exact point in that interval is objectively ambiguous to player 2 and herself. It is as-if the player uses an Ellsberg urn – an ambiguous randomization device – to decide on the action to take.

- Player 1's set of *Ellsberg strategies* is $E_1 := \{[p_1, p_2] \mid 0 \leq p_1 \leq p_2 \leq 1\}$ with generic element $e_1 := [p_1, p_2]$. Analogously, for player 2 we have $E_2 := \{[q_1, q_2] \mid 0 \leq q_1 \leq q_2 \leq 1\}$ with generic element $e_2 := [q_1, q_2]$.

  Observe that mixed strategies belong to the set of Ellsberg strategies. The Ellsberg strategy $[p_1, p_2]$ is a *proper* Ellsberg strategy if the interval is non-degenerate: $p_1 < p_2$.

- Let the pair $e := (e_1, e_2) = \big([p_1, p_2], [q_1, q_2]\big) \in E_1 \times E_2$ denote an Ellsberg strategy profile.

The decision making of agents confronted with ambiguous outcomes depends on their attitudes with respect to ambiguity. Some empirical evidence summarized in Camerer and Weber (1992) suggests that agents are ambiguity averse, i.e. agents are pessimistic in the face of multiple priors. In decision theory, Gilboa and Schmeidler (1989) have shown that ambiguity averse agents evaluate ambiguous outcomes by considering the worst point in their set of priors. Their decision rule is therefore of the maxmin type. More recently, Gajdos et al. (2008) have axiomatized the minimal expected utility evaluation of ambiguous outcomes for strategic settings.

We assume players are ambiguity averse. Given the result of Gilboa and Schmeidler (1989), players expected utility is therefore computed based on the worst point in the interval.

- Player 1's expected utility for the strategy profile (e_1, e_2) is:

  $$U_1(e_1, e_2) = \min_{p \in e_1, q \in e_2} U_1(p, q)$$

  By the linearity of $U_1(p, q)$, we have:

  $$U_1\big(p, [q_1, q_2]\big) = \min\Big(U_1(p, q_1), U_1(p, q_2)\Big) \tag{4.1}$$

- For player 1, strategy e_1 is a best reply to e_2 if $U_1(e_1, e_2) \geq U_1(e_1', e_2)$ for all $e_1' \in E_1$.
- Strategy profile $e = (e_1, e_2)$ is an *Ellsberg equilibrium* if e_1 and e_2 are mutual best replies.
- The equilibrium $e = (e_1, e_2)$ is a *proper* Ellsberg equilibrium if both equilibrium strategies are proper Ellsberg strategies. It is a *quasi-proper* Ellsberg equilibrium if only one equilibrium strategy is a proper Ellsberg strategy and the other strategy is a mixed strategy.

4.2.2 The class of games considered

We restrict our attention to 2×2 normal form games satisfying two restrictions. First, we assume that no player has a weakly dominant strategy. As shown by Harsanyi (1973), games with weakly dominant strategies admit Nash equilibria that cannot be purified. Discarding *weakly dominant strategies* rules out games that are Row Dominant for player 1 and games that are Column Dominant for player 2.

Definition 29 (Row Dominant).
Player i's payoffs in Γ *are row dominant if* $(\pi_i^1 - \pi_i^3)(\pi_i^2 - \pi_i^4) \geq 0$.

Definition 30 (Column Dominant).
Player i's payoffs in Γ *are column dominant if* $(\pi_i^1 - \pi_i^2)(\pi_i^3 - \pi_i^4) \geq 0$.

The introduction of the second restriction requires some additional definitions. Riedel and Sass (2013) show that two types of mixed strategies play an important role for (quasi-) proper Ellsberg equilibria. These strategies are central in our disambiguation result.

Definition 31 (Indifference Strategy).
Strategy p^* *is an indifference strategy for player 1 if:*

$$U_2(q, p^*) = U_2(q', p^*) \text{ for all } q, q' \in [0, 1].$$

Strategy q^* *is an indifference strategy for player 2 if:*

$$U_1(p, q^*) = U_1(p', q^*) \text{ for all } p, p' \in [0, 1].$$

In words, playing your indifference strategy makes your opponent indifferent between all her mixed strategies q. By definition, the pair (p^*, q^*) constitutes a Nash equilibrium in mixed strategies. As shown in Lemma 22, all games satisfying *No weakly dominant strategy* have a unique equilibrium in proper mixed strategies. Therefore, indifference strategies p^* and q^* exist and are unique. Next, we define maxmin strategies.

Definition 32 (Maxmin Strategy).
Strategy $\bar{p}$ *is a maxmin strategy for player 1 if:*

$$\bar{p} = arg \max_{p \in [0,1]} \min_{q \in [0,1]} U_1(p, q).$$

Strategy $\bar{q}$ is a maxmin strategy for player 2 if:

$$\bar{q} = arg \max_{q \in [0,1]} \min_{p \in [0,1]} U_2(q, p).$$

In words, playing your maxmin strategy guarantees you the highest payoff if your opponent aims at minimizing your payoff and anticipates your strategy correctly. As shown by von Neumann and Morgenstern, maxmin strategies exist in 2×2 games – they are unique for the games we consider – and the maxmin stategy coincides with the indifference strategy in zero-sum games.

The maxmin strategy is a proper mixed strategy for the subset of games characterized in statements 2 and 3 of Lemma 22. For these games, a player using her maxmin strategy is "immunized" against her opponent's strategy. This implies that her maxmin strategy makes her indifferent between all her opponent's strategies. Such a strategy therefore yields a safe expected payoff.

Lemma 22.
Any game Γ with no weakly dominant strategies has the following properties.

1. *Indifference strategies p^* and q^* are unique, maxmin strategies $\bar{p}$ and $\bar{q}$ are unique and $p^*, q^* \in (0,1)$.*

2. *If player 1's payoffs are* not *Column Dominant in game Γ, then $\bar{p} \in (0,1)$ and $U_1(\bar{p}, q) = U_1(\bar{p}, q')$ for all $q, q' \in [0,1]$.*

3. *If player 2's payoffs are* not *Row Dominant in game Γ, then $\bar{q} \in (0,1)$ and $U_2(\bar{q}, p) = U_2(\bar{q}, p')$ for all $p, p' \in [0,1]$.*

Proof. See Appendix 4.6.1. ■

Riedel and Sass (2013) show that for games in which indifference and maxmin strategies coincide, a particular type of Ellsberg equilibria arises for which the indifference strategy belongs to the interior of the Ellsberg strategy. This type of Ellsberg equilibria can not be disambiguated. This should not be seen as a problem however because these equilibria are non-robust.[1] Our second restriction rules out these games.[2]

Definition 33 (Class of games $\mathbf{\Gamma}$)**.**
A 2×2 normal form game Γ belongs to the class $\mathbf{\Gamma}$ if no player has a weakly dominant strategy and for each player, the indifference and maxmin strategies do not coincide.

As we show in Lemma 23, for all (quasi-) proper Ellsberg equilibria of games in $\mathbf{\Gamma}$ and for each player, the indifference strategy lies at an extreme point of the equilibrium Ellsberg strategy.

[1] Slight perturbations to the payoffs destroy these equilibria.

[2] In terms of payoffs, $p^* \neq \bar{p}$ is equivalent to $\frac{\pi_2^4 - \pi_2^3}{\pi_2^4 - \pi_2^3 + \pi_2^1 - \pi_2^2} \neq \frac{\pi_1^4 - \pi_1^3}{\pi_1^4 - \pi_1^3 + \pi_1^1 - \pi_1^2}$ and $q^* \neq \bar{q}$ is equivalent to $\frac{\pi_1^4 - \pi_1^2}{\pi_1^4 - \pi_1^2 + \pi_1^1 - \pi_1^3} \neq \frac{\pi_2^4 - \pi_2^2}{\pi_2^4 - \pi_2^2 + \pi_2^1 - \pi_2^3}$.

Γ_1		q L	1-q R
p	U	$4,4$	$0,3$
1-p	D	$3,0$	$2,2$

Γ_2		q L	1-q R
p	U	$2,0$	$0,1$
1-p	D	$0,1$	$1,0$

Figure 4.2: Γ_1 is a game of class I for which $p^* = q^* = \frac{2}{3}$ and $\bar{p} = \bar{q} = 0$. Γ_2 is a game of class II for which $p^* = \frac{1}{2}$, $\bar{p} = \frac{1}{3}$, $q^* = \frac{1}{3}$ and $\bar{q} = \frac{1}{2}$.

Lemma 23.
For all $\Gamma \in \mathbf{\Gamma}$, if $\big([p_1, p_2], [q_1, q_2]\big)$ is a (quasi-) proper Ellsberg equilibrium, then $p^* \in \{p_1, p_2\}$ and $q^* \in \{q_1, q_2\}$.

Proof. See Appendix 4.6.2. ■

The interpretation of our Disambiguation Theorem is different depending on the class to which the game belongs. We divide our family of games into two classes I and II, which are illustrated in Figure 4.2.

Definition 34 (Row and column dominance)**.**
Consider any $\Gamma \in \mathbf{\Gamma}$. If player 1's payoffs are Column Dominant and player 2's payoffs are Row Dominant, then Γ belongs to class I, otherwise Γ belongs to class II.

4.2.3 The disturbed games $\Gamma^*(\epsilon)$

For any basic game $\Gamma \in \mathbf{\Gamma}$, we define a parametric family of disturbed games whose generic member $\Gamma^*(\epsilon)$ is shown in Figure 4.3. Payoffs in $\Gamma^*(\epsilon)$ are affected by the realization of ambiguous disturbances. The size of the disturbances is parameterized by $\epsilon \geq 0$. When ϵ is zero, the disturbed game is equivalent to the basic game. The ambiguous random variables r and t are private information of player 1 and 2 respectively. Their common support is $[-1, 1]$. We emphasize that the disturbances in Harsanyi (1973) are payoff-specific, which is not the case in our framework. For simplicity, we require the disturbance to be strategy-specific: the payoffs of outcomes associated to the same pure strategy are subject to identical disturbances. As disturbances are strategy-specific, they enter the evaluation of strategy profiles as an additional term independent of the opponent's strategy:

$$U_1\big(p, [q_1, q_2], \epsilon r\big) = U_1\big(p, [q_1, q_2]\big) + p\epsilon r, \tag{4.2}$$
$$U_2\big(q, [p_1, p_2], \epsilon t\big) = U_2\big(q, [p_1, p_2]\big) + q\epsilon t. \tag{4.3}$$

Observe that when maxmin strategies yield a safe payoff in the basic game, they keep this property in the disturbed games:

$$U_1\big(\bar{p}, [q_1, q_2], \epsilon r\big) = U_1\big(\bar{p}, [q'_1, q'_2], \epsilon r\big) \text{ for all } [q_1, q_2], [q'_1, q'_2] \in E_2,$$
$$U_2\big(\bar{q}, [p_1, p_2], \epsilon t\big) = U_2\big(\bar{q}, [p'_1, p'_2], \epsilon t\big) \text{ for all } [p_1, p_2], [p'_1, p'_2] \in E_1.$$

			Player 2	
			q	1-q
			L	R
Player 1	p	U	$\pi_1^1 + \epsilon \mathbf{r}$; $\pi_2^1 + \epsilon \mathbf{t}$	$\pi_1^2 + \epsilon \mathbf{r}; \pi_2^2$
	1-p	D	$\pi_1^3; \pi_2^3 + \epsilon \mathbf{t}$	$\pi_1^4; \pi_2^4$

Figure 4.3: The normal form of the disturbed game $\Gamma^*(\epsilon)$ associated with the basic game Γ. The realization of the disturbances r and t are the private information of player 1 and 2, respectively.

Unlike in Harsanyi (1973), the density f_r of the random variable r over her support $[-1, 1]$ is unknown. Players only have partial information about the density f_r. They only know the domain $\mathcal{P}_r$ of f_r. The domain $\mathcal{P}_r$ summarizes all of the information available to players about the density f_r. We define $\mathcal{P}_r$ to be a ball in the set of densities around a known basic density f_r^b. The basic density f_r^b belongs to the set $\mathcal{F}$ of measurable densities with full support on $[-1, 1]$.

$$\mathcal{P}_r := \left\{ f \in \mathcal{F} \,\middle|\, f_r^b(x)(1-k_r) \leq f(x) \leq f_r^b(x)(1+k_r) \text{ for all } x \in [-1,1] \right\}.$$

The parameter $k_r \in [0, 1]$ can be interpreted as the radius of the ball since k_r defines the maximal deviation from the basic density. It measures the level of ambiguity associated with the domain $\mathcal{P}_r$. When $k_r = 0$, density f_r is known – $f_r = f_r^b$ – and there is no ambiguity. The ambiguity is maximal for $k_r = 1$. At this value, not all elements $f \in \mathcal{P}_r$ have full support. This way of defining a domain from a basic density is a form of *ϵ-contamination*, as defined in the literature on ambiguous variables (see Huber, P. (1981), Eichberger and Kelsey (2000) or Maccheroni et al. (2006)). Analogously, the random variable t has unknown density $f_t \in \mathcal{P}_t$ and $\mathcal{P}_t$ is characterized by the basic density f_t^b and the ambiguity parameter $k_t \in [0, 1]$.

Strategies in the disturbed game $\Gamma^*(\epsilon)$ are functions from the space of possible realizations of the disturbances to the set of mixed strategies.

- Let $p^b : [-1, 1] \to [0, 1]$ be a generic strategy for player 1 in the disturbed game. For player 1, the set of strategies in the disturbed game is denoted by S_1 and contains only measurable functions p^b. Analogously, a generic strategy for player 2 in the disturbed game is $q^b \in S_2$.

How do players perceive the strategy of their opponent in the disturbed game? Suppose player 2 anticipates correctly the strategy p^b of player 1. Player 2 ignores the realization of r but knows the domain $\mathcal{P}_r$ in which f_r lies. For each density $f \in \mathcal{P}_r$ of the random variable r, strategy p^b implies that pure strategy U is played with a probability p. The probability of playing U is minimal for the density in $\mathcal{P}_r$ that puts maximal weight on the realization of r for which strategy p^b prescribes low values of p. Let this minimal probability be denoted by p_{min}. Conversely, a maximal probability p_{max} is implied by the density in $\mathcal{P}_r$ that puts maximal weight on the realization of r for which strategy p^b prescribes high values of p. As Lemma 24 shows, all probabilities $p \in [p_{min}, p_{max}]$ result from some density in $\mathcal{P}_r$. Therefore, when player 2 believes that player 1 plays p^b, player 2 anticipates that the probability that player 1 uses strategy U lies in some interval of probabilities $[p_{min}, p_{max}]$. In other words, player 2 perceives player 1's strategy as the Ellsberg strategy $[p_{min}, p_{max}]$. In our terminology, this Ellsberg strategy is *induced* by the strategy p^b. Observe that if $k_r = 0$, then $p_{min} = p_{max}$ and the induced strategy is a mixed strategy.

Lemma 24.
Any strategy $p^b \in S_1$ induces an Ellsberg strategy $\left[p_{min}, p_{max}\right] \subseteq [0,1]$ defined by:

$$p_{min} = \min_{f \in \mathcal{P}_r} \int_{-1}^{1} p^b(r) f(r) dr \qquad \text{and} \qquad p_{max} = \max_{f \in \mathcal{P}_r} \int_{-1}^{1} p^b(r) f(r) dr. \tag{4.4}$$

Proof. See Appendix 4.6.3. ■

Equivalently, any strategy $q^b \in S_2$ induces an Ellsberg strategy $\left[q_{min}, q_{max}\right]$. A direct consequence of Lemma 24 and equations (4.2) and (4.3) is the following: for player 2, the expected utility of playing q when player 1 uses the strategy p^b inducing $\left[p_{min}, p_{max}\right]$ is given by:

$$U_2(q, p^b, \epsilon t) = U_2\big(q, [p_{min}, p_{max}]\big) + q\epsilon t. \tag{4.5}$$

The equivalent equation for player 1 is:

$$U_1(p, q^b, \epsilon r) = U_1\big(p, [q_{min}, q_{max}]\big) + p\epsilon r. \tag{4.6}$$

For brevity, we often refer to strategies p^b and q^b by the Ellsberg strategies they induce, respectively $\left[p_{min}, p_{max}\right]$ and $\left[q_{min}, q_{max}\right]$.

We now define best replies and equilibria in the disturbed games.

- Strategy p^b is a best reply to strategy q^b inducing $\left[q_{min}, q_{max}\right]$ if we have $U_1(p^b, \left[q_{min}, q_{max}\right], \epsilon r) \geq U_1(p^{b'}, \left[q_{min}, q_{max}\right], \epsilon r)$ for all $r \in [-1,1]$ and all $p^{b'} \in S_1$.
- The profile (p^b, q^b) is an equilibrium in the disturbed game $\Gamma^*(\epsilon)$ if p^b and q^b are mutual best-replies. The corresponding induced Ellsberg equilibrium is written $e(\epsilon) = \big([p_{min}, p_{max}], [q_{min}, q_{max}]\big)$.

Two categories of strategies in disturbed games are focal best replies, namely pure and maxmin strategies. These strategies are monotone in the realization of the ambiguous variable and are based on threshold values for r and t.

Definition 35 (Pure and maxmin strategies in a disturbed game)**.**
*The strategy p^b_{pu} is a **pure strategy** in $S^1_{pu} \subset S_1$ if there exists a single threshold $r^* \in \mathbb{R}$ such that:*

$$p^b_{pu}(r) = \begin{cases} 0 & \text{if } r \leq r^*, \\ 1 & \text{if } r > r^*. \end{cases}$$

The strategy p^b_{mm} *is a* ***maxmin strategy*** *in* $S^1_{mm} \subset S_1$ *if there exist two thresholds* $r_1, r_2 \in \mathbb{R}$ *such that:*

$$p^b_{mm}(r) = \begin{cases} 0 & \text{if } r < r_1, \\ \bar{p} & \text{if } r_1 \leq r \leq r_2, \\ 1 & \text{if } r > r_2. \end{cases}$$

In a disturbed game, pure strategies are a special case of maxmin strategies. If the maxmin strategy $\bar{p}$ is a pure strategy of the basic game, then a maxmin strategy in the disturbed game is pure. Pure and maxmin strategies in the disturbed game for player 2 are defined accordingly. For brevity, we refer to maxmin strategies p^b_{mm} or q^b_{mm} by their two thresholds (r_1, r_2) or (t_1, t_2).

4.3 The Disambiguation Theorem

This section presents and proves a disambiguation theorem for 2×2 normal form games in $\mathbf{\Gamma}$. This theorem is the central result of this paper. The interpretation that our result gives to Ellsberg equilibria is contained in the definition of purifiable and disambiguable equilibria.

Definition 36 (Purifiable and disambiguable equilibria).
Let $e = \big([p_1, p_2], [q_1, q_2]\big)$ *be an Ellsberg equilibrium in* $\Gamma \in \mathbf{\Gamma}$.

- *Equilibrium* e *is* ***purifiable*** *if for some pair* $(k_r, k_t) \in [0,1] \times [0,1]$, *there exists a sequence of* pure strategy *equilibria in* $\Gamma^*(\epsilon)$ *inducing outcomes* $e(\epsilon)$ *with*
$$\lim_{\epsilon \to 0} e(\epsilon) = e.$$
- *Equilibrium* e *is* ***disambiguable*** *if for some pair* $(k_r, k_t) \in [0,1] \times [0,1]$, *there exists a sequence of* maxmin strategy *equilibria in* $\Gamma^*(\epsilon)$ *inducing outcomes* $e(\epsilon)$ *with*
$$\lim_{\epsilon \to 0} e(\epsilon) = e.$$

Notice that purifiable equilibria are a subset of disambiguable equilibria as pure strategies in the disturbed games are a subset of maxmin strategies.

Theorem 11 (Disambiguation of Ellsberg equilibria).
All (quasi-) proper Ellsberg equilibria in games of class I are purifiable.
All (quasi-) proper Ellsberg equilibria in games of class II are disambiguable.

In the remainder of this section, we present a proof of Theorem 11. The proof often requires considering different cases. For clarity, we focus on the following subset of games.

Definition 37 (Subset $\mathbf{\Gamma}^{II-D}$ of games of class II).
Let $\mathbf{\Gamma}^{II-D} \subset \mathbf{\Gamma}$ *be the subset of games for which player 1's payoffs are* not *Column Dominant and player 2's payoffs are* not *Row Dominant.*

For these games, maxmin strategies $\bar{p}$ and $\bar{q}$ are proper mixed strategies. We focus on these games because they are the most difficult case and best illustrate the consequences of ambiguity. At the end of this section, we discuss the small adaptations needed to extend the proof to other types of games in $\mathbf{\Gamma}$.

The proof is structured as follows. First, we provide sufficient conditions for a strategy profile to be an equilibrium in a disturbed game. Second, we show how these sufficient conditions simplify for small disturbances. For small disturbances, there is a unique threshold per best reply that lies in the support $[-1, 1]$. Third, we prove the existence of equilibria in disturbed games for small disturbances. Then, abstracting from equilibrium conditions, we show that, for small disturbances, any Ellsberg strategy potentially involved in an Ellsberg equilibrium of the basic game can be induced by a unique value of the ambiguity parameter. Finally, we bring all of these findings together to prove the theorem.

4.3.1 Sufficient conditions for an equilibrium in $\Gamma^*(\epsilon)$

In Harsanyi's Purification Theorem, best replies to the realization of the disturbances turn out to be in pure strategies. This needs not be the case in our setting for which the induced strategies are Ellsberg strategies. For games in $\mathbf{\Gamma}^{II-D}$, on top of their pure strategies, players best reply using their maxmin strategies $\bar{p}$ and $\bar{q}$. Best replies of player 1 are monotone in r and make use of her pure strategies and her maxmin strategy. For all games for which player 1's payoffs are *not* Column Dominant, Lemma 25 provides conditions under which a maxmin strategy is a best reply to a strategy of player 2 inducing $[q_{min}, q_{max}]$.

Lemma 25 (Best-Reply in maxmin strategies)**.**
*For all $\epsilon > 0$ and all $\Gamma \in \mathbf{\Gamma}$ such that player 1's payoffs are **not** Column Dominant, strategy p^b is a best reply to any $[q_{min}, q_{max}] \subseteq [0, 1]$ if it is a maxmin strategy $p^b = (r_1, r_2) \in S^1_{mm}$ defined by:*

$$\begin{aligned} \epsilon r' =& U_1(0, q_{min}) - U_1(1, q_{min}), \\ \epsilon r'' =& U_1(0, q_{max}) - U_1(1, q_{max}), \\ r_1 =& min(r', r''), \\ r_2 =& max(r', r''). \end{aligned}$$

Proof. Take any $\epsilon > 0$, any $\Gamma \in \mathbf{\Gamma}$ such that player 1's payoffs are not Column Dominant and any $[q_{min}, q_{max}] \subseteq [0, 1]$. Given equations (4.1) and (4.2), we have

$$\begin{aligned} U_1\big(p, [q_{min}, q_{max}], \epsilon r\big) &= min\big(U_1(p, q_{min}, \epsilon r), U_1(p, q_{max}, \epsilon r)\big) \\ &= min\big(U_1(p, q_{min}) + p\epsilon r, U_1(p, q_{max}) + p\epsilon r\big) \end{aligned}$$

where $U_1(p, q, \epsilon r)$ is linear in p since $U_1(p, q)$ is linear in p. Let $q^1 := q_{min}$ and $q^2 := q_{max}$ if $r_1 = r'$ and $q^1 := q_{max}$ and $q^2 := q_{min}$ otherwise. By definition of r_1 and r_2 we have

$$U_1(0, q^1) + 0\epsilon r_1 = U_1(1, q^1) + 1\epsilon r_1 \quad \text{and} \quad U_1(0, q^2) + 0\epsilon r_2 = U_1(1, q^2) + 1\epsilon r_2.$$

Remembering that $r_1 \leq r_2$, these definitions imply

- $U_1(1, q^1) + 1\epsilon r < U_1(0, q^1) + 0\epsilon r$ for all $r < r_1$,
- $U_1(1, q^2) + 1\epsilon r < U_1(0, q^2) + 0\epsilon r$ for all $r < r_1$.

The last two inequalities imply that for all $r < r_1$, both $U_1(p, q^1, \epsilon r)$ and $U_1(p, q^2, \epsilon r)$ are strictly decreasing in p because they both are linear in p. Therefore the unique best reply when $r < r_1$ is to take $p = 0$. The same definitions also imply that

- $U_1(1, q^1) + 1\epsilon r > U_1(0, q^1) + 0\epsilon r$ for all $r > r_2$,
- $U_1(1, q^2) + 1\epsilon r > U_1(0, q^2) + 0\epsilon r$ for all $r > r_2$.

The last two inequalities imply that for all $r > r_2$, both $U_1(p, q^1, \epsilon r)$ and $U_1(p, q^2, \epsilon r)$ are strictly increasing in p. Therefore the unique best reply when $r > r_2$ is to take $p = 1$. Finally we have

- $U_1(1, q^1) + 1\epsilon r < U_1(0, q^1) + 0\epsilon r$ for all r with $r_1 < r < r_2$,
- $U_1(1, q^2) + 1\epsilon r > U_1(0, q^2) + 0\epsilon r$ for all r with $r_1 < r < r_2$.

For those intermediate values of r, $U_1(p, q^1, \epsilon r)$ is strictly decreasing in p while $U_1(p, q^2, \epsilon r)$ is strictly increasing in p. By definition of $\bar{p}$, $U_1(p, q^1, \epsilon r)$ and $U_1(p, q^2, \epsilon r)$ cross in $p = \bar{p}$. By Lemma 22, $\bar{p}$ is unique and belongs to $(0, 1)$ since player 1's payoffs are not Column Dominant. The unique best reply is to take $p = \bar{p}$. Finally, when $r = r_1$ or $r = r_2$, either $U_1(p, q_{min}, \epsilon r)$ or $U_1(p, q_{max}, \epsilon r)$ is constant in p. A (non-unique) best reply is then $p = \bar{p}$. Notice that this proof also covers the case $q_{min} = q_{max}$. ■

For all games for which player 2's payoffs are not Row Dominant, parallel conditions guarantee that a maxmin strategy of player 2 is a best reply to a strategy of player 1 inducing $[p_{min}, p_{max}]$.

Thresholds r_1 and r_2 defined above belong to $\mathbb{R}$. The exact values taken by those thresholds matter for p^b_{mm} only as long as they belong to the support $[-1, 1]$. For example, $(r_1, r_2) = (0, 2)$ induces the same reactions to the disturbances realization as $(r'_1, r'_2) = (0, 4)$, since 2 and 4 do not belong to the support.

Lemma 26 provides sufficient conditions for the strategy profile $\big((r_1, r_2), (t_1, t_2)\big)$ to be an equilibrium in the disturbed game.

Lemma 26 (Sufficient conditions for equilibrium in $\Gamma^*(\epsilon)$)**.**
For all $\epsilon > 0$, $\Gamma \in \mathbf{\Gamma}^{II-D}$ and $(k_r, k_t) \in [0, 1] \times [0, 1]$, the profile of maxmin strategies $\big((r_1, r_2), (t_1, t_2)\big) \in S^1_{mm} \times S^2_{mm}$ is an equilibrium in $\Gamma^*(\epsilon)$ if equations (4.7) to (4.14) hold:[3]

$$\epsilon r' = U_1(0, q_{min}) - U_1(1, q_{min}), \qquad \epsilon r'' = U_1(0, q_{max}) - U_1(1, q_{max}), \tag{4.7}$$

$$\epsilon t' = U_2(0, p_{min}) - U_2(1, p_{min}), \qquad \epsilon t'' = U_2(0, p_{max}) - U_2(1, p_{max}), \tag{4.8}$$

[3]Equations (4.11) to (4.14) correspond to the case for which all thresholds belong to the support. If it was not the case, the expression for these integrals should be modified. Any threshold outside the support must be replaced by the nearest point in the support. These modifications are necessary for equations (4.11) to (4.14) to correspond to equation (4.4).

$$r_1 = min(r', r''), \qquad r_2 = max(r', r''), \tag{4.9}$$
$$t_1 = min(t', t''), \qquad t_2 = max(t', t''), \tag{4.10}$$

$$p_{min} = \min_{f \in \mathcal{P}_r} \int_{-1}^{r_1} 0 f(r)dr + \int_{r_1}^{r_2} \bar{p} f(r)dr + \int_{r_2}^{1} 1 f(r)dr, \tag{4.11}$$

$$p_{max} = \max_{f \in \mathcal{P}_r} \int_{-1}^{r_1} 0 f(r)dr + \int_{r_1}^{r_2} \bar{p} f(r)dr + \int_{r_2}^{1} 1 f(r)dr, \tag{4.12}$$

$$q_{min} = \min_{f \in \mathcal{P}_t} \int_{-1}^{t_1} 0 f(t)dt + \int_{t_1}^{t_2} \bar{q} f(t)dt + \int_{t_2}^{1} 1 f(t)dt, \tag{4.13}$$

$$q_{max} = \max_{f \in \mathcal{P}_t} \int_{-1}^{t_1} 0 f(t)dt + \int_{t_1}^{t_2} \bar{q} f(t)dt + \int_{t_2}^{1} 1 f(t)dt. \tag{4.14}$$

Proof. Take any $\epsilon > 0$, any $\Gamma \in \mathbf{\Gamma}^{II-D}$, any $(k_r, k_t) \in [0,1] \times [0,1]$ and any profile of maxmin strategies $\big((r_1, r_2), (t_1, t_2)\big) \in S^1_{mm} \times S^2_{mm}$ for which equations (4.7) to (4.14) hold. From Lemma 24, the extreme points of the induced Ellsberg strategies $[p_{min}, p_{max}]$ and $[q_{min}, q_{max}]$ of any profile $\big((r_1, r_2), (t_1, t_2)\big)$ are given by equations (4.11) to (4.14). As $\Gamma \in \mathbf{\Gamma}^{II-D}$, we have that player 1's payoffs are not Column Dominant. From Lemma 25, the best-reply of player 1 to $[q_{min}, q_{max}]$ is to use a strategy (r_1, r_2) whose thresholds r_1 and r_2 are defined by equations (4.7) and (4.9). Accordingly, the best-reply for player 2 to $[p_{min}, p_{max}]$ is a maxmin strategy (t_1, t_2), whose thresholds t_1 and t_2 are defined by equations (4.8) and (4.10). Therefore, if all equations hold, strategies (r_1, r_2) and (t_1, t_2) are mutual best replies and the profile constitutes an equilibrium in $\Gamma^*(\epsilon)$. ■

4.3.2 Simplified conditions for small disturbances

In this subsection we show how the previous conditions simplify when the size of disturances ϵ is sufficiently small. These simpler conditions are given in Lemma 30. Intermediary lemmas and definitions are necessary for proving Lemma 30. Lemmas 27 to 29 study the conditions on induced Ellsberg strategies under which the thresholds generated by the strategies lie in the support. The lemmas also identify some properties of the induced Ellsberg strategies when the thresholds lie in the support. First, Lemma 27 describes the interval of probabilities in which the extreme points of the induced Ellsberg strategy must lie in order for their associated threshold to be in the support. New notations are necessary for establishing this lemma:

- The equilibrium conditions given in Lemma 26 link thresholds t_1 and t_2 of player 2's strategy to the two extreme points of player 1's induced Ellsberg strategy $[p_{min}, p_{max}]$. Hence there exists an interval $\big[p^-(\epsilon), p^+(\epsilon)\big] \subset \mathbb{R}$, inside which p_{min} and p_{max} must lie in order for their associated thresholds t_1 and t_2

to be in $[-1,1]$.[4]

$$p^{-}(\epsilon) := min\left\{p \in \mathbb{R} \;\middle|\; \frac{1}{\epsilon}\big(U_2(0,p) - U_2(1,p)\big) \in [-1,1]\right\},$$
$$p^{+}(\epsilon) := max\left\{p \in \mathbb{R} \;\middle|\; \frac{1}{\epsilon}\big(U_2(0,p) - U_2(1,p)\big) \in [-1,1]\right\}.$$

Similarly, we define for player 2 the interval $\big[q^{-}(\epsilon), q^{+}(\epsilon)\big]$.

$$q^{-}(\epsilon) := min\left\{q \in \mathbb{R} \;\middle|\; \frac{1}{\epsilon}\big(U_1(0,q) - U_1(1,q)\big) \in [-1,1]\right\},$$
$$q^{+}(\epsilon) := max\left\{q \in \mathbb{R} \;\middle|\; \frac{1}{\epsilon}\big(U_1(0,q) - U_1(1,q)\big) \in [-1,1]\right\}.$$

Lemma 27 shows that for all ϵ, the indifference strategy p^* lies in the interval $\big(p^{-}(\epsilon), p^{+}(\epsilon)\big)$ and q^* lies in $\big(q^{-}(\epsilon), q^{+}(\epsilon)\big)$. Furthermore, those intervals collapse on p^* and q^* when $\epsilon \to 0$.

Lemma 27.
For all $\epsilon > 0$ and $\Gamma \in \mathbf{\Gamma}$ we have $p^{-}(\epsilon) < p^* < p^{+}(\epsilon)$ and $q^{-}(\epsilon) < q^* < q^{+}(\epsilon)$. Furthermore:

$$\lim_{\epsilon \to 0} p^{-}(\epsilon) = \lim_{\epsilon \to 0} p^{+}(\epsilon) = p^* \qquad \textit{and} \qquad \lim_{\epsilon \to 0} q^{-}(\epsilon) = \lim_{\epsilon \to 0} q^{+}(\epsilon) = q^*.$$

Proof. Take any $\epsilon > 0$ and any $\Gamma \in \mathbf{\Gamma}$. We focus on proving this for the interval $\big(p^{-}(\epsilon), p^{+}(\epsilon)\big)$, the reasoning is identical for $\big(q^{+}(\epsilon), q^{-}(\epsilon)\big)$. The expression

$$\frac{1}{\epsilon}\big(U_2(0,p) - U_2(1,p)\big)$$

returning the thresholds on t is linear in p. As $\Gamma \in \mathbf{\Gamma}$, player 2 does not have a weakly dominant strategy and therefore this expression is strictly monotone in p. Therefore $p^{-}(\epsilon)$ and $p^{+}(\epsilon)$ are finite and hence exist. By definition of p^*, this expression equal 0 for $p = p^*$. Therefore we have $p^{-}(\epsilon) < p^* < p^{+}(\epsilon)$ by the strict monotonicity of the above linear expression.

The difference $U_2(0,p) - U_2(1,p)$ is independent of ϵ. As a result, for any $p \neq p^*$ the smaller ϵ, the larger $\left|\frac{1}{\epsilon}\big(U_2(0,p) - U_2(1,p)\big)\right|$. Hence, for any $p \neq p^*$, there exists an ϵ^p such that for all $\epsilon < \epsilon^p$, we have $\frac{1}{\epsilon}\big(U_2(0,p) - U_2(1,p)\big) \notin [-1,1]$. Therefore, the smaller ϵ, the closer $p^{-}(\epsilon)$ and $p^{+}(\epsilon)$ are to p^*. In the limit, the interval $[p^{-}(\epsilon), p^{+}(\epsilon)]$ collapse on p^*. ■

Lemmas 28 and 29 provide bounds around the extreme points of the induced Ellsberg strategies when thresholds lie in the interior of the support and ambiguity is strictly positive. If the domain $\mathcal{P}_r$ contains a strictly positive amount of ambiguity, Lemma 28 shows that the induced Ellsberg strategy cannot degenerate into a mixed strategy when a threshold lies in the interior of the support. If one extreme point of the induced Ellsberg strategy for player 1 equals p, the other extreme point lies outside a non-degenerate interval (p^{ℓ}, p^{u}) around p. New notations are necessary for establishing Lemma 28.

[4] As shown in the proof of Lemma 27, the extreme points of the interval $\big[p^{-}(\epsilon), p^{+}(\epsilon)\big]$ exist.

- As shown in Lemma 25, player 1 best replies to any realization of r by playing a strategy in $\{0, \bar{p}, 1\}$. Equivalently, player 2's best reply to t lies in $\{0, \bar{q}, 1\}$. For small disturbances, at most two of the three strategies in those sets are used in equilibrium. In the absence of ambiguity, only pure strategies are used. In the presence of ambiguity, the payoff structure determines for each player which two strategies among these three strategies are used in equilibrium. These two strategies are referred to as A and B for player 1 and C and D for player 2. Let $A, B \in \{0, \bar{p}, 1\}$ with $A < B$ and $C, D \in \{0, \bar{q}, 1\}$ with $C < D$ be such that:

$$\begin{array}{lll} A := 0 \text{ and } B := 1 & if\ k_t = 0, \\ A := 0 \text{ and } B := \bar{p} & if\ k_t > 0 \text{ and } p^* \in (0, \bar{p}), \\ A := \bar{p} \text{ and } B := 1 & if\ k_t > 0 \text{ and } p^* \in (\bar{p}, 1), \\ C := 0 \text{ and } D := 1 & if\ k_r = 0, \\ C := 0 \text{ and } D := \bar{q} & if\ k_r > 0 \text{ and } q^* \in (0, \bar{q}), \\ C := \bar{q} \text{ and } D := 1 & if\ k_r > 0 \text{ and } q^* \in (\bar{q}, 1). \end{array}$$

  The maxmin strategy of one player is never used when the ambiguity about her opponent's payoffs is zero. For example, if $k_r = 0$ then the induced Ellsberg strategy of player 1 is a mixed strategy and player 2 best replies using pure strategies.

- Let function $p_{min} : \mathbb{R}^2 \to [0, 1] : (r_1, r_2) \to p_{min}(r_1, r_2)$ be defined by equation (4.11). Accordingly, functions p_{max}, q_{min} and q_{max} are defined by equations (4.12), (4.13) and (4.14) respectively.

Lemma 28.
For all $\Gamma \in \mathbf{\Gamma}^{II-D}$, $k_r \in (0, 1]$ and $p \in (A, B)$, there exist unique p^ℓ and $p^u \in [0, 1]$ with $p^\ell < p < p^u$ such that

- *for all $(r_1, r_2) \in S^1_{mm}$ with $p_{max}(r_1, r_2) = p$ we have $p_{min}(r_1, r_2) \leq p^\ell$; and at least for one such (r_1, r_2) we have $p_{min}(r_1, r_2) = p^\ell$,*

- *for all $(r_1, r_2) \in S^1_{mm}$ with $p_{min}(r_1, r_2) = p$, $p_{max}(r_1, r_2) \geq p^u$, and at least for one such (r_1, r_2) we have $p_{max}(r_1, r_2) = p^u$,*

Accordingly, for all $k_t \in (0, 1]$ and $q \in (C, D)$, there exist $q^\ell, q^u \in [0, 1]$ with equivalent properties.

Proof. Take any $\Gamma \in \mathbf{\Gamma}^{II-D}$, any $k_r \in (0, 1]$ and any $p \in (A, B)$. We focus on proving the existence of such p^ℓ and p^u. The proof for q^ℓ and q^u follows the same reasoning. We define the following sets:

$$S^{max}(p) := \left\{(r_1, r_2) \in S^1_{mm} \,\middle|\, p_{max}(r_1, r_2) = p \text{ and } r_1, r_2 \in [-1, 1]\right\},$$
$$S^{min}(p) := \left\{(r_1, r_2) \in S^1_{mm} \,\middle|\, p_{min}(r_1, r_2) = p \text{ and } r_1, r_2 \in [-1, 1]\right\}.$$

$S^{max}(p)$ is a subset of the maxmin strategies whose induced Ellsberg strategies have

p as their maximal point. We show below this set is non-empty. The restriction $r_1, r_2 \in [-1, 1]$ implies that $S^{max}(p)$ and $S^{min}(p)$ are closed sets.

We define p^ℓ from the set $S^{max}(p)$:

$$p^\ell := \max_{(r_1,r_2)\in S^{max}(p)} p_{min}(r_1, r_2).$$

As the domain of images of function p_{min} is $[0, 1]$ and the set $S^{max}(p)$ is non-empty and closed, p^ℓ is well defined. The definitions of p^ℓ and $S^{max}(p)$ imply that p^ℓ is such that:

(i) for all $(r_1, r_2) \in S^{max}(p)$ we have $p_{max}(r_1, r_2) = p$ and $p_{min}(r_1, r_2) \leq p^\ell$; and at least for one such (r_1, r_2) we have $p_{min}(r_1, r_2) = p^\ell$, and

(ii) there is no $p' \neq p^\ell$ with the previous properties.

We next show that $p^\ell < p$. As $\Gamma \in \mathbf{\Gamma}^{II-D}$ we have $\bar{p} \in (A, B)$, and hence two cases can arise:

- Case 1: $p^* > \bar{p}$. This case is such that $A = \bar{p}$ and $B = 1$ and by assumption we have $p \in (\bar{p}, 1)$. Let r_2^L and r_2^H be implicitly defined by

  $$p_{max}(-1, r_2^H) = p \qquad \text{and} \qquad p_{max}(r_2^L, r_2^L) = p.$$

  We show that for all $(r_1, r_2) \in S^{max}(p)$, we have $-1 < r_2^L \leq r_2 \leq r_2^H < 1$. Observe this implies that $S^{max}(p)$ is a non-empty set.

  - First we show $-1 < r_2^L < 1$.

    For all $k_r \in (0, 1]$, because of its integral functional form, the expression of $p_{max}(x, x)$ is continuous in x. Furthermore, it is decreasing in x for $x \in [-1, 1)$ as maxmin strategies are increasing in r. Since $p_{max}(-1, -1) = 1$, $p_{max}(1, 1) = 0$ and by assumption $p \in (\bar{p}, 1)$, we therefore have $-1 < r_2^L < 1$.

  - Second we show $-1 < r_2^H < 1$.

    For all $k_r \in (0, 1]$, the expression of $p_{max}(-1, x)$ is continuous in x and decreasing in x for $x \in [-1, 1)$. Since $p_{max}(-1, -1) = 1$, $p_{max}(-1, 1) = \bar{p}$ and by assumption $p \in (\bar{p}, 1)$, we therefore have $-1 < r_2^H < 1$.

  - Then we show that $r_2^L < r_2^H$.

    Assume instead that $r_2^L \geq r_2^H$. As by definition $p_{max}(-1, r_2^H) = p$, we have $p_{max}(r_2^H, r_2^H) < p$ as for all $k_r \in (0, 1]$ and $r_1, r_2 \in [-1, 1)$, p_{max} is a strictly decreasing function of both r_1 and r_2 and we showed $-1 < r_2^H$. As $r_2^L \geq r_2^H$ the same reasoning implies $p_{max}(r_2^L, r_2^L) \leq p_{max}(r_2^H, r_2^H) < p$, contradiction the definition of r_2^L.

  - Finally we show that for all $(r_1, r_2) \in S^{max}(p)$ we have $r_2^L \leq r_2 \leq r_2^H$.

    We focus on showing $r_2 \leq r_2^H$, the proof that $r_2^L \leq r_2$ follows similar lines. Assume instead for some $(r_1, r_2) \in S^{max}(p)$ that $r_2 > r_2^H$. By the definition of $S^{max}(p)$ we have $-1 \leq r_1$. As p_{max} is strictly decreasing in its argument, this implies that $p_{max}(r_1, r_2) \leq p_{max}(-1, r_2)$. As we assumed $r_2 > r_2^H$, the same reasoning implies $p_{max}(-1, r_2) < p_{max}(-1, r_2^H) = p$. Together we have $p_{max}(r_1, r_2) < p$, implying that $(r_1, r_2) \notin S^{max}(p)$, a contradiction.

- Case 2: $p^* < \bar{p}$. This second case is such that $A = 0$ and $B = \bar{p}$ and by assumption we have $p \in (0, \bar{p})$. Let r_1^L and r_1^H be implicitly defined by $p_{max}(r_1^H, 1) = p$ and $p_{max}(r_1^L, r_1^L) = p$. The proof showing that for all $(r_1, r_2) \in S^{max}(p)$, we have $-1 < r_1^L \leq r_1 \leq r_1^H < 1$ is omitted as it follows the lines of that given for case 1.

Together, either there exist r_1^L and r_1^H such that for all $(r_1, r_2) \in S^{max}(p)$ we have $-1 < r_1^L \leq r_1 \leq r_1^H < 1$ or there exist r_2^L and r_2^H such that for all $(r_1, r_2) \in S^{max}(p)$ we have $-1 < r_2^L \leq r_2 \leq r_2^H < 1$. This implies $min(\mid r_1 \mid, \mid r_2 \mid) < 1$.

From there, as $k_r > 0$ we have for all $(r_1, r_2) \in S^{max}(p)$ that

$$p_{min}(r_1, r_2) < p_{max}(r_1, r_2)$$

because

(i) $p_{min}(r_1, r_2) = p_{max}(r_1, r_2)$ when $k_r = 0$ and,

(ii) for all (r_1, r_2) with $min(\mid r_1 \mid, \mid r_2 \mid) < 1$, p_{max} is a strictly increasing function of k_r at all $k_r \in [0, 1)$ while p_{min} is a strictly decreasing function of k_r.

This proves that $p^\ell < p$.

There remains to show that p^ℓ has the same properties for all $(r_1, r_2) \in S^1_{mm}$. As the support of r is $[-1, 1]$, for any $(r_1, r_2) \in S^1_{mm}$ with $p_{max}(r_1, r_2) = p$ such that $(r_1, r_2) \notin S^{max}(p)$, there exists $(r'_1, r'_2) \in S^{max}(p)$ inducing the same Ellsberg strategy as (r_1, r_2). Therefore p^ℓ has the desired properties.

We define then p^u from the set $S^{min}(p)$:

$$p^u := \min_{(r_1, r_2) \in S^{min}(p)} p_{max}(r_1, r_2).$$

An analog reasoning proves that p^u has the desired properties. ■

Lemma 29 shows that the interval (p^ℓ, p^u) around p defined in the previous lemma evolves monotonically with p.

Lemma 29.
Take any $\Gamma \in \mathbf{\Gamma}^{II-D}$.

- *For all* $k_r \in (0, 1)$, $p \in (A, B)$ *and* $p' \in (p^\ell, p)$ *we have*
$$p'^\ell < p^\ell < p' < p < p'^u < p^u.$$

- *For all* $k_t \in (0, 1)$, $q \in (C, D)$ *and* $q' \in (q^\ell, q)$ *we have*
$$q'^\ell < q^\ell < q' < q < q'^u < q^u.$$

Table 4.1: Four types of (quasi)-proper Ellsberg equilibria $\big([p_1, p_2], [q_1, q_2]\big)$ in function of the extreme point occupied by the indifference strategy of each player. For quasi-proper Ellsberg equilibria, the extreme points of one player are equal and those equilibria belong to two of the above-defined types. The mixed strategy equilibrium (p^*, q^*) belongs to all four.

	$q^* = q_1$	$q^* = q_2$
$p^* = p_1$	type 1	type 2
$p^* = p_2$	type 3	type 4

Proof. We focus on proving the first claim. The proof is based on the properties of functions p_{min} and p_{max}. Those functions are continuous in both their arguments r_1 and r_2. Furthermore, they are non-increasing in both arguments and strictly decreasing as soon as these arguments belong to $[-1, 1)$.

Take any $\Gamma \in \mathbf{\Gamma}^{II-D}$, any $k_r \in (0, 1)$, any $p \in (A, B)$ and any $p' \in (p^\ell, p)$. From Lemma 28, we have that $p'^\ell < p^\ell < p'^u$. We show by contradiction that $p'^u < p^u$, $p < p'^u$ and $p'^\ell < p^\ell$.

Assume first $p^u \leq p'^u$. This implies by definition of p'^u that there does not exist (r_1, r_2) with $p_{min}(r_1, r_2) = p'$ and $p_{max}(r_1, r_2) < p^u$. Take (r_1', r_2') with $p_{min}(r_1', r_2') = p$ and $p_{max}(r_1', r_2') = p^u$. By definition of p^u, this (r_1', r_2') exists and has at least one threshold in the interior of the support. By continuity and non-increasingness of p_{min}, there exists (r_1, r_2) with $r_1 > r_1'$ and $r_2 > r_2'$ such that $p_{min}(r_1, r_2) = p'$. Since $p \in (A, B)$, we have either $r_1' \in (-1, 1)$ or $r_2' \in (-1, 1)$.[5] By the properties of p_{max}, we have $p_{max}(r_1, r_2) < p_{max}(r_1', r_2') = p^u$, a contradiction.

Assume then that $p'^u \leq p$. This implies by definition of p'^u that there exists (r_1, r_2) with $p_{min}(r_1, r_2) = p'$ and $p_{max}(r_1, r_2) \leq p$. By continuity and non-increasingness of p_{max}, there exists (r_1', r_2') with $r_1' \leq r_1$ and $r_2' \leq r_2$ such that $p_{max}(r_1', r_2') = p$. By the properties of p_{min}, we have $p_{min}(r_1', r_2') \geq p_{min}(r_1, r_2) = p'$, a contradiction to the definition of p^ℓ since $p^\ell < p'$.

Assume finally that $p^\ell \leq p'^\ell$. This implies by definition of p'^ℓ that there exists (r_1, r_2) with $p_{max}(r_1, r_2) = p'$ and $p_{min}(r_1, r_2) \geq p^\ell$. By continuity and non-increasingness of p_{max}, there exists (r_1', r_2') with $r_1' < r_1$ and $r_2' < r_2$ such that $p_{max}(r_1', r_2') = p$. By the properties of p_{min}, we have $p_{min}(r_1', r_2') > p_{min}(r_1, r_2) \geq p^\ell$, which contradicts the definition of p^ℓ. ■

When only one of the two thresholds lies in the interior of the support, we denote this threshold r^* for player 1 and t^* for player 2. The equilibrium conditions simplify. Nevertheless, their expressions will depend on the type of equilibrium we consider. Lemma 23 shows that for both players the indifference strategy is an extreme point of Ellsberg strategies in any (quasi)-proper Ellsberg equilibrium. The expression of the conditions depends on whether this extreme point is the maximum or the minimum. The four different types of (quasi-) proper Ellsberg equilibria are presented in Table 4.1.

[5] See proof of Lemma 28.

Riedel and Sass (2013) present results linking the payoff structure with the extreme points occupied by the indifference strategy of each player. As all proper Ellsberg equilibria of a game belong to the same type, these types naturally define subsets of games. We denote $\mathbf{\Gamma}^{D-4} \subset \mathbf{\Gamma}^{II-D}$ the subset of games in $\mathbf{\Gamma}^{II-D}$ having proper Ellsberg equilibria of type 4. For the rest of the proof, we concentrate exclusively on equilibria of type 4, which are illustrated in the example developed in section 4.4. The proof presented is easily adapted for the other types.

In Lemma 30, we give simplified sufficient conditions for a profile of maxmin strategies to constitute an equilibrium in slightly disturbed games.

Lemma 30 (Simplified equilibrium conditions in slightly disturbed games)**.** *For all* $\Gamma \in \mathbf{\Gamma}^{D-4}$ *and* $(k_r, k_t) \in [0,1] \times [0,1]$*, there exists* $\bar{\epsilon} > 0$ *such that for all* $\epsilon < \bar{\epsilon}$*, if the profile of maxmin strategies* $\big((r_1, r_2), (t_1, t_2)\big)$ *satisfies conditions* (4.15) *to* (4.18) *and equations* (4.19) *to* (4.22)*, then it is an equilibrium in* $\Gamma^*(\epsilon)$*.*

$$\underline{\text{if } p^* < \bar{p}}: \quad r^* := \min(r_1, r_2) \in [-1,1] \quad \text{and if } r_1 < r_2: \quad \max(r_1, r_2) \geq 1, \tag{4.15}$$
$$\underline{\text{if } p^* > \bar{p}}: \quad r^* := \max(r_1, r_2) \in [-1,1] \quad \text{and if } r_1 < r_2: \min(r_1, r_2) \leq -1, \tag{4.16}$$
$$\underline{\text{if } q^* < \bar{q}}: \quad t^* := \min(t_1, t_2) \in [-1,1] \quad \text{and if } t_1 < t_2: \quad \max(t_1, t_2) \geq 1, \tag{4.17}$$
$$\underline{\text{if } q^* > \bar{q}}: \quad t^* := \max(t_1, t_2) \in [-1,1] \quad \text{and if } t_1 < t_2: \quad \min(t_1, t_2) \leq -1, \tag{4.18}$$
$$\epsilon r^* = U_1(0, q_{max}) - U_1(1, q_{max}), \tag{4.19}$$
$$\epsilon t^* = U_2(0, p_{max}) - U_2(1, p_{max}), \tag{4.20}$$
$$p_{max} = \max_{f \in \mathcal{P}_r} \int_{-1}^{r^*} A f(r) dr + \int_{r^*}^{1} B f(r) dr, \tag{4.21}$$
$$q_{max} = \max_{f \in \mathcal{P}_t} \int_{-1}^{t^*} C f(t) dt + \int_{t^*}^{1} D f(t) dt. \tag{4.22}$$

Proof. We show that such maxmin strategies (r_1, r_2) and (t_1, t_2) are mutual best replies. Take any $\Gamma \in \mathbf{\Gamma}^{D-4}$. Lemma 26 gives sufficient conditions for such profile to be an equilibrium. In these conditions, the following additional four equations complement equations (4.19) to (4.22):

$$\epsilon r' = U_1(0, q_{min}) - U_1(1, q_{min}),$$
$$\epsilon t' = U_2(0, p_{min}) - U_2(1, p_{min}),$$
$$p_{min} = \min_{f \in \mathcal{P}_r} \int_{-1}^{r^*} A f(r) dr + \int_{r^*}^{1} B f(r) dr,$$
$$q_{min} = \min_{f \in \mathcal{P}_t} \int_{-1}^{t^*} C f(t) dt + \int_{t^*}^{1} D f(t) dt.$$

We show there exists $\bar{\epsilon} > 0$ such that for all $\epsilon < \bar{\epsilon}$, if $\big((r_1, r_2), (t_1, t_2)\big)$ satisfy conditions (4.15) to (4.18), then $r' \notin [-1,1]$ and $t' \notin [-1,1]$ and hence those four additional equations are irrelevant for the profile to be an equilibrium.

By definition of A and B we have $p^* \in (A, B)$. By Lemma 27, there exists $\bar{\epsilon}_1 > 0$ such that for all $\epsilon < \bar{\epsilon}_1$ we have $\big[p^-(\epsilon), p^+(\epsilon)\big] \subset (A, B)$. Accordingly, we have $q^* \in (C, D)$ and there exists $\bar{\epsilon}_2 > 0$ such that for all $\epsilon < \bar{\epsilon}_2$ we have

$\big[q^-(\epsilon), q^+(\epsilon)\big] \subset (C, D)$. As $\big((r_1, r_2), (t_1, t_2)\big)$ satisfies conditions (4.15) to (4.18) we have $r^* \in [-1, 1]$ and $t^* \in [-1, 1]$ and hence two cases must be considered.

Case 1: $|r^*| = 1$ or $|t^*| = 1$.

Taking $\bar{\epsilon} = min(\bar{\epsilon}_1, \bar{\epsilon}_2)$ we derive a contradiction for this case. Assume that $|t^*| = 1$. Conditions (4.17) and (4.18) imply that $min(| t_1 |, | t_2 |) \geq 1$. The maxmin strategy (t_1, t_2) is such that $q^b(t)$ is the same for all $t \in [-1, 1]$ with $q^b(t) \in \{C, D\}$ and therefore $q_{min} = q_{max} \in \{C, D\}$. As for all $\epsilon < \bar{\epsilon}$ we have $\big[q^-(\epsilon), q^+(\epsilon)\big] \subset (C, D)$, this implies that either $r^* \notin [-1, 1]$, which violates condition (4.15) or (4.16), or equation (4.19) does not hold.

Case 2: $|r^*| < 1$ or $|t^*| < 1$.

Proving that (r_1, r_2) and (t_1, t_2) are mutual best replies boils down to showing that

(i) r' and t' are not in the support and,

(ii) the relative size of r^* and r' makes it optimal for player 1 to react to r using strategies A and B, as well as it is optimal for player 2 to react to t using C and D given the relative size of t^* and t'.

If (i) and (ii) hold, then equations (4.19) to (4.22) are a simplification of equations (4.7) to (4.14) and the strategies are mutual best replies. Two subcases must be considered

- Subcase 2.1: $k_r > 0$ and $k_t > 0$.

  The profile of maxmin strategies $\big((r_1, r_2), (t_1, t_2)\big)$ induces *proper* Ellsberg strategies since thresholds r^* and t^* lie in the interior of $[-1, 1]$.[6] Player 1's proper Ellsberg strategy has two different extreme points p_{max} and p_{min} which induce two different thresholds t^* and t' for player 2. Accordingly we have $q_{max} \neq p_{min}$ and hence $t^* \neq t'$.

  We show here (i), that is r' and t' are not in the support. Given $k_r > 0$ and $k_t > 0$, by Lemma 28 and Lemma 29 there exist p_L, $p_U \in (A, B)$ and q_L, $q_U \in (C, D)$ such that

$$p^\ell_L \leq p^\ell_U < p_L < p^* < p_U < p^u_L \leq p^u_U,$$
$$q^\ell_L \leq q^\ell_U < q_L < q^* < q_U < q^u_L \leq q^u_U.$$

  such that if $p_{max} \in [p_L, p_U]$, then $p_{min} \notin [p_L, p_U]$ and if $q_{max} \in [q_L, q_U]$, then $q_{min} \notin [q_L, q_U]$. We prove the existence of such p_L and p_U. As $p^* \in (A, B)$, given $k_r > 0$, Lemma 28 shows there exists $p^{*\ell}$ and p^{*u} with $p^{*\ell} < p^* < p^{*u}$

[6] This statement holds as well when $k_r = 1$ or $k_t = 1$ as $p_{max} \in (A, B)$ and $q_{max} \in (C, D)$, implying respectively that $p_{min} \in \{A, B\}$ or $q_{min} \in \{C, D\}$.

such that if $p_{max} = p^*$, then $p_{min} \leq p^{*\ell}$ and if $p_{min} = p^*$, then $p_{max} \geq p^{*u}$. Take any $p_L \in (p^{*\ell}, p^*)$. By Lemma 29, we have $p^* < p_L^u$. Take p_U such that $p^* < p_U < p_L^u$. By Lemma 29 we have $p_L^\ell < p_U^\ell < p_L < p^* < p_U < p_L^u < p_U^u$, hence the desired property for $[p_L, p_U]$.[7]

Let $\bar{\epsilon}' > 0$ be such that $\bar{\epsilon}' \leq \bar{\epsilon}_1$ and for all $\epsilon < \bar{\epsilon}'$ we have $\left[p^-(\epsilon), p^+(\epsilon)\right] \subset (p_L, p_U)$. By Lemma 27, this $\bar{\epsilon}'$ exists since $\left[p^-(\epsilon), p^+(\epsilon)\right]$ tends to $[p^*, p^*]$ as $\epsilon \to 0$. The same reasoning proves the existence of an $\bar{\epsilon}'' > 0$ such that $\bar{\epsilon}'' \leq \bar{\epsilon}_2$ and for all $\epsilon < \bar{\epsilon}''$ we have $\left[q^-(\epsilon), q^+(\epsilon)\right] \subset (q_L, q_U)$. Take $\bar{\epsilon} = min(\bar{\epsilon}', \bar{\epsilon}'')$.

By the construction of $\bar{\epsilon}$, for all $\epsilon < \bar{\epsilon}$ conditions (4.15) and (4.16) combined with equation (4.19) imply that $q_{max} \in \left[q^-(\epsilon), q^+(\epsilon)\right] \subset [q_L, q_U]$ and hence $q_{min} \notin [q_L, q_U]$, therefore $q_{min} \notin \left[q^-(\epsilon), q^+(\epsilon)\right]$, implying $r' \notin [-1, 1]$. A parallel reasoning shows $t' \notin [-1, 1]$.

We turn to proving (ii). We focus on showing that the relative sizes of t^* and t' make it optimal for player 2 to react to t using strategies C and D. A parallel argument demonstrates that player 1 best replies using A and B. As $\Gamma \in \mathbf{\Gamma}^{D-4}$ we have $\bar{p} \in (A, B)$, and hence two subcases can arise:

- Subcase 2.1.1: $\bar{q} < q^*$.

  Assume for a moment that the difference $U_2(0, p) - U_2(1, p)$ is a strictly *increasing* function of p. As $p_{min} < p_{max}$, this assumption implies that for a given ϵ we have $t' < t^*$ and hence $t_1 = t'$ and $t_2 = t^*$. As $t^* \in [-1, 1]$ and $t' \notin [-1, 1]$, we have $t_1 < -1$. By definition, when $\bar{q} < q^*$, we have $C = \bar{q}$ and $D = 1$. It is hence optimal for player 2 to react to the realization of t using C and D, as shown in the proof of Lemma 25.

  There remains to show that the difference $U_2(0, p) - U_2(1, p)$ is a strictly *increasing* function of p. The difference $U_2(0, p) - U_2(1, p)$ is linear in p and can not be constant since weakly dominant strategies are ruled out. By definition, any game $\Gamma \in \mathbf{\Gamma}^{D-4}$ has proper Ellsberg equilibria $e = \big([p_1, p_2], [q_1, q_2]\big)$ of type 4, for which $p_2 = p^*$ and hence $p_1 < p^*$. In order for $[q_1, q_2]$ to be a best reply to $[p_1, p_2]$, we must have $q^* \in \{q_1, q_2\}$ as shown in Lemma 23. For $q^* \in \{q_1, q_2\}$, we must have

  $$U_2(q, p_1) > U_2(q, p^*) \text{ for all } q \in (\bar{q}, 1].$$

  In effect, remember that the definition of p^* implies that $U_2(0, p^*) - U_2(1, p^*) = 0$ and hence $U_2(q, p^*)$ is constant in q. The definition of $\bar{q}$ implies $U_2(\bar{q}, p^*) = U_2(\bar{q}, p_1)$. If we had instead for all $q \in (\bar{q}, 1]$ that $U_2(q, p_1) < U_2(q, p^*)$, then $U_2(q, p_1)$ is strictly decreasing in q and the best reply for the ambiguity averse player 2 to $[p_1, p^*]$ would be some $[q_1, q_2] \subset [0, \bar{q}]$, contradicting Lemma 23 since $\bar{q} < q^*$. As $U_2(\bar{q}, p^*) = U_2(\bar{q}, p_1)$ and $U_2(1, p_1) > U_2(1, p^*)$, we have $U_2(0, p_1) < U_2(0, p^*)$. Last two inequalities imply $U_2(0, p_1) - U_2(1, p_1) < U_2(0, p^*) - U_2(1, p^*)$ and

[7] Weak inequalities $p_L^\ell \leq p_U^\ell$ and $p_L^u \leq p_U^u$ come from the case $k_r = 1$. For such value of k_r, we have $p_L^\ell = p_U^\ell = A$ and $p_L^u = p_U^u = B$, as shown in the proof of Lemma 33.

by linearity of $U_2(q,p)$ in p, the difference $U_2(0,p) - U_2(1,p)$ is a strictly *increasing* function of p as $p_1 < p^*$.

– Subcase 2.1.2: $q^* < \bar{q}$.

The argument follow the same line as for the previous case. The major difference is that $U_2(0,p) - U_2(1,p)$ must now be strictly *decreasing* function of p. As $p_{min} < p_{max}$, this implies that for a given ϵ we have $t^* < t'$ and hence $t_1 = t^*$ and $t_2 = t'$. As $t^* \in [-1,1]$ and $t' \notin [-1,1]$, we have $t_2 > 1$. By definition, when $\bar{q} < q^*$, we have $C = 0$ and $D = \bar{q}$. It is hence optimal for player 2 to react to the realization of t using C and D.

Statement (ii) holds as for each of the above subcases, conditions (4.17) and (4.18) pick t^* among t_1 and t_2 consistently with the particular game considered and ensure that t' is outside the support with the appropriate relative size with respect to t^*.

- Subcase 2.1: $k_r = 0$ or $k_t = 0$.

  We consider only $k_r = 0$, without loss of generality. This implies that $C = 0$ and $D = 1$, $p_{min} = p_{max}$ and $t' = t^*$. Both t' and t^* belong to the interior of the support. The induced Ellsberg profile is quasi-proper. Except for these differences, the argument given above to prove (i) and (ii) carries on to this subcase.

■

4.3.3 Existence of equilibria

Showing existence of equilibria in the disturbed game is much easier for small disturbances. The simplified conditions of Lemma 30 are such that only one threshold per strategy is constrained. The other threshold can be picked arbitrarily provided it lies outside the support and has the appropriate sign.

Lemma 31 (Existence of equilibria in disturbed games)**.**
For all $\Gamma \in \mathbf{\Gamma}^{D-4}$ and $(k_r, k_t) \in [0,1] \times [0,1]$, there exists $\overline{\epsilon} > 0$ such that for all $\epsilon < \overline{\epsilon}$, equilibria exist in $\Gamma^*(\epsilon)$.

Proof. Using the Intermediate Value Theorem, we show the existence of a profile of thresholds (r^*, t^*) satisfying equations (4.19) to (4.22). If it exists, then it is easy to see that there always exists a strategy profile $\big((r_1, r_2), (t_1, t_2)\big)$ that, together with (r^*, t^*), satisfies conditions (4.15) to (4.18). Such strategy profile $\big((r_1, r_2), (t_1, t_2)\big)$ satisfies the conditions of Lemma 30 for small ϵ. This proves the existence of equilibria in slightly disturbed games.

There remains to show the existence of a profile of thresholds (r^*, t^*) satisfying equations (4.19) to (4.22). Take any $\Gamma \in \mathbf{\Gamma}^{D-4}$ and any $(k_r, k_t) \in [0,1] \times [0,1]$. We define the four functions h_r, h_t, h_p and h_q:

$$\begin{aligned}
h_r :&[q^-(\epsilon), q^+(\epsilon)] \to [-1,1] : q \to h_r(q) := \frac{1}{\epsilon}\big(U_1(0,q) - U_1(1,q)\big),\\
h_t :&[p^-(\epsilon), p^+(\epsilon)] \to [-1,1] : p \to h_t(p) := \frac{1}{\epsilon}\big(U_2(0,p) - U_2(1,p)\big),\\
h_p :&[-1,1] \to [A,B] : r \to h_p(r) := \max_{f\in\mathcal{P}_r} \int_{-1}^{r} Af(r)dr + \int_{r}^{1} Bf(r)dr,\\
h_q :&[-1,1] \to [C,D] : t \to h_q(t) := \max_{f\in\mathcal{P}_t} \int_{-1}^{t} Cf(t)dt + \int_{t}^{1} Df(t)dt.
\end{aligned}$$

Those four functions are all strictly monotone and continuous. By the definition of $p^-(\epsilon)$, $p^+(\epsilon)$, $q^-(\epsilon)$ and $q^+(\epsilon)$, the domain of images of h_r and h_t is $[-1,1]$ and hence all four functions are surjective. The strict monotonicity of these functions imply they are injective. Being all bijective (surjective and injective), they admit inverse functions h_r^{-1}, h_t^{-1}, h_p^{-1}, h_q^{-1} which are strictly monotone and continuous.[8]

Based on these four functions, we define two composite functions g_1 and g_2:

$$\begin{aligned}
g_1 :&[A,B] \to [q^-(\epsilon), q^+(\epsilon)] : p \to g_1(p) := h_r^{-1} \circ h_p^{-1}(p) = (h_p \circ h_r)^{-1}(p),\\
g_2 :&[C,D] \to [p^-(\epsilon), p^+(\epsilon)] : q \to g_2(q) := h_t^{-1} \circ h_q^{-1}(q) = (h_q \circ h_t)^{-1}(q).
\end{aligned}$$

Being composite functions of strictly monotone and continuous functions, g_1 and g_2 inherit those properties.

By Lemma 27, there exists $\bar{\epsilon}' > 0$ such that for all $\epsilon < \bar{\epsilon}'$ we have $[p^-(\epsilon), p^+(\epsilon)] \subset (A,B)$ and $[q^-(\epsilon), q^+(\epsilon)] \subset (C,D)$. Those two composite functions are then used to define the continuous mapping τ:

$$\tau : [A,B] \to [p^-(\epsilon), p^+(\epsilon)] : p \to \tau(p) := g_2 \circ g_1(p).$$

We have therefore that for all $\epsilon < \bar{\epsilon}'$, τ is a continuous mapping from $[A,B] \to [p^-(\epsilon), p^+(\epsilon)] \subset (A,B)$. By the Intermediary Value Theorem, it has a fixed point $\hat{p} \in [p^-(\epsilon), p^+(\epsilon)]$. This fixed point is associated to $\hat{q} = g_1(\hat{p})$ as well as $\hat{r} = h_r(\hat{q})$ and $\hat{t} = h_t(\hat{p})$. By construction, these $\hat{r}$, $\hat{t}$, $\hat{p}$ and $\hat{q}$ satisfy equations (4.19) to (4.22) in Lemma 30. Let $\bar{\epsilon}''$ be taken from the statement of Lemma 30. Taking $\bar{\epsilon} = min(\bar{\epsilon}', \bar{\epsilon}'')$ completes the proof. ■

4.3.4 The limit of the sequence of equilibria

There remains to prove that, when the disturbance size vanishes, the Ellsberg equilibrium induced in the disturbed game tends to the equilibrium in the initial game. By Lemma 27 and Lemma 30, slightly disturbed games admit equilibria inducing

[8] The strict monotonicity of h_p and h_q is only valid as long as $k_r < 1$ and $k_t < 1$. When $k_r = 1$ ($k_t = 1$), function h_p (h_q) is not injective. This is not a problem for our purpose as these functions are injective and surjective on a smaller domain. For example, when $k_r = 1$, function h_p is bijective on $[\hat{r}, 1] \subset [-1,1]$ defined in the proof of Lemma 33. The definition of probabilities $p^-(\epsilon)$ and $p^+(\epsilon)$ must be adapted such that h_r has the appropriate domain of image $[\hat{r}, 1]$. On this basis a similar mapping can be constructed.

Ellsberg strategies with an extreme point close to the indifference strategy. Furthermore, as the disturbance size vanishes, the extreme point tends to the indifference strategy. We show in Lemma 32 that, for each particular level of ambiguity, a unique threshold's value induces such an extreme point. Then, given the other extreme point of the equilibrium Ellsberg strategy in the basic game, Lemma 33 shows there exists an appropriate level of ambiguity for the induced Ellsberg strategy to reproduce the equilibrium in the basic game. More precisely, abstracting from equilibrium conditions, any Ellsberg strategy can be induced for a unique value of the ambiguity parameter. We introduce new notations that are useful for small disturbances:

$$p_{min}(r^*) \coloneqq \min_{f \in \mathcal{P}_r} \int_{-1}^{r^*} Af(r)dr + \int_{r^*}^{1} Bf(r)dr, \tag{4.23}$$

$$p_{max}(r^*) \coloneqq \max_{f \in \mathcal{P}_r} \int_{-1}^{r^*} Af(r)dr + \int_{r^*}^{1} Bf(r)dr, \tag{4.24}$$

$$q_{min}(t^*) \coloneqq \min_{f \in \mathcal{P}_t} \int_{-1}^{t^*} Cf(t)dt + \int_{t^*}^{1} Df(t)dt, \tag{4.25}$$

$$q_{max}(t^*) \coloneqq \max_{f \in \mathcal{P}_t} \int_{-1}^{t^*} Cf(t)dt + \int_{t^*}^{1} Df(t)dt. \tag{4.26}$$

Abstracting from equilibrium conditions, Lemma 32 shows that for any given k_r, there exists a unique threshold r^* for which the induced Ellsberg strategy has the desired value for one of the two extreme points.

Lemma 32 (Uniqueness of r^*).
Consider any $\Gamma \in \mathbf{\Gamma}$ and $(k_r, k_t) \in [0,1] \times [0,1]$.

- *For all $p \in (A, B)$, there is a unique $r^* \in (-1, 1)$ such that $p_{min}(r^*) = p$.*
- *For all $p \in (A, B)$, there is a unique $r^{*'} \in (-1, 1)$ such that $p_{max}(r^{*'}) = p$.*

For all $q \in (C, D)$, equivalent t^* and $t^{*'}$ are also unique.

Proof. We prove only the existence and uniqueness of r^*. Function p_{min} is continuous and weakly decreasing in r^*.[9] Furthemore, $p_{min}(-1) = B$ and $p_{min}(1) = A$. By continuity, there exists hence $r^* \in (-1, 1)$ such that $p_{min}(r^*) = p$. We prove now uniqueness.

For all $k_r \in [0, 1)$, all $f \in \mathcal{P}_r$ have full support. Function p_{min} is therefore strictly decreasing for all $r \in [-1, 1]$, which entails uniqueness of r^*.

For the case $k_r = 1$, let $R_A \coloneqq \{r \in [-1, 1] | p_{min}(r) = A\}$ and let $\hat{r} \coloneqq \min\{r \in R_A\}$. This $\hat{r}$ exists since the set R_A is a non-degenerate closed interval. Uniqueness of r^* is ensured since $p > A$ and for all $r \in [-1, \hat{r})$, function p_{min} is strictly decreasing in r. ∎

Lemma 33 (All equilibrium Ellsberg strategies can be induced).
Consider any $\Gamma \in \mathbf{\Gamma}$.

[9] Function p_{min} is *strictly* decreasing in r^* when $k_r < 1$.

1. *For all $p_1 \in (A, B)$ and all $p_2 \in [p_1, B]$, there exists a unique $k_r \in [0, 1]$ such that for some $r^* \in [-1, 1]$ we have $p_{min}(r^*) = p_1$ and $p_{max}(r^*) = p_2$.*

2. *For all $p_2 \in (A, B)$ and all $p_1 \in [A, p_2]$, there exists a unique $k'_r \in [0, 1]$ such that for some $r^{*\prime} \in [-1, 1]$ we have $p_{min}(r^{*\prime}) = p_1$ and $p_{max}(r^{*\prime}) = p_2$.*

Equivalent statements hold true for player 2.

Proof. We prove only the first of the two claims.
By Lemma 32, for all $p_1 \in (A, B)$ and all $k_r \in [0, 1]$, there exists a unique $r^* \in (-1, 1)$ such that $p_{min}(r^*) = p_1$. Let $F : [0, 1] \to [-1, 1] : k_r \to F(k_r)$ be the function pointing, for each value of $k_r \in [0, 1]$, to the particular r^* inducing $p_{min}(r^*) = p_1$ and hence $F(k_r) = r^*$. From equation (4.23), function F is continuous and strictly decreasing in k_r as $p_1 \in (A, B)$.

As F is continuous, the composite function $p_{max} \circ F : [0, 1] \to [p_1, b] : k_r \to p_{max}\big(F(k_r)\big)$ is continuous and strictly increasing in k_r as $p_1 \in (A, B)$. The first claim we need to prove follows then from the fact that $p_{max}\big(F(0)\big) = p_{min}\big(F(0)\big) = p_1$ and $p_{max}\big(F(1)\big) = B$.

The equality $p_{max}\big(F(1)\big) = B$ follows from the definition of the domain $\mathcal{P}_r$. For $k_r = 1$, some $f \in \mathcal{P}_r$ do not have full support anymore and there exists a unique $\hat{r} \in (-1, 1)$ such that:

$$\max_{f \in \mathcal{P}_r} \int_{-1}^{\hat{r}} f(r)dr = 1 \qquad \text{and} \qquad \min_{f \in \mathcal{P}_r} \int_{-1}^{\hat{r}} f(r)dr = 0.$$

This $\hat{r}$ is implicitly defined by $\int_{-1}^{\hat{r}} f_r^b(r)dr = \frac{1}{2}$. As a result, if $p_{min} = p_1 < 1$, then $r^* < \hat{r}$ and $p_{max} = B$. In effect, when $k_r = 1$, for all $r \in (-1, 1)$ either $p_{min}(r) = A$ or $p_{max}(r) = B$. ■

We emphasize again that we did not require in Lemmas 32 and 33 that threshold r^* corresponds to any kind of best reply. Moreover the previous lemmas holds true independently of the value taken by ϵ.

4.3.5 Proof of the theorem

We rephrase here Theorem 11 for $\Gamma \in \mathbf{\Gamma}^{D-4}$. The proof given covers a particular subset of games but similar proofs can easily be constructed to extend the proof to the full family of games $\mathbf{\Gamma}$.

Theorem 1 (Disambiguation of equilibria in $\Gamma \in \mathbf{\Gamma}^{D-4}$).
For all $\Gamma \in \mathbf{\Gamma}^{D-4}$ and all (quasi-) proper Ellsberg equilibrium $e = \big([p_1, p_2], [q_1, q_2]\big)$ in Γ, there exists a unique pair $(k_r, k_t) \in [0, 1] \times [0, 1]$ for which there exists a sequence of Ellsberg strategy profiles $\{e(\epsilon)\}$ induced by equilibria in $\Gamma^*(\epsilon)$ with

$$\lim_{\epsilon \to 0} e(\epsilon) = e.$$

Proof. Take any $\Gamma \in \mathbf{\Gamma}^{D-4}$. Let $\{\Gamma^*(\epsilon^t)\}_{t=1}^{\infty}$ be a sequence of disturbed games for which $\epsilon^t := \frac{1}{t}$ for all $t \in \mathbb{N}$. This sequence is defined such that $\lim_{t \to \infty} \Gamma^*(\epsilon^t) = \Gamma$. By

Lemma 31, for all $(k_r, k_t) \in [0,1]\times[0,1]$ there exists $\overline{\epsilon} > 0$ such that for all $\epsilon < \overline{\epsilon}$ there exist equilibria in maxmin strategies in the disturbed game. Let T be the smallest $t \in \mathbb{N}$ such that $\epsilon^t < \overline{\epsilon}$. There exists hence a sequence of equilibria $\{e(\epsilon^t)\}_{t=T}^{\infty}$, one for each disturbed game in $\{\Gamma^*(\epsilon^t)\}_{t=T}^{\infty}$, for which $\epsilon^t \to 0$ when $t \to \infty$. From now on, this sequence is denoted $\{e(\epsilon)\}$. There remains to show that for all (quasi-) proper Ellsberg equilibrium e in Γ, there exists $(k_r, k_t) \in [0,1] \times [0,1]$ such that this sequence has the right limit.

By the definition of $\mathbf{\Gamma}^{D-4}$, we have for all (quasi-) proper Ellsberg equilibria $e = \big([p_1, p_2], [q_1, q_2]\big)$ that $p_2 = p^*$ and $q_2 = q^*$. Riedel and Sass (2013) have shown for such games that e is a (quasi-) proper Ellsberg equilibrium if and only if $A \leq p_1 \leq p^*$ and $C \leq q_1 \leq p^*$. Consider any p_1 and q_1 satisfying those constraints.

Take any (quasi-) proper Ellsberg equilibria $e = \big([p_1, p_2], [q_1, q_2]\big)$ for game Γ. Let $k_r \in [0,1]$ be such that for some $r^* \in [-1,1]$, we have $p_{min}(r^*) = p_1$ and $p_{max}(r^*) = p_2$. Let $k_t \in [0,1]$ be such that for some $t^* \in [-1,1]$, we have $q_{min}(t^*) = q_1$ and $q_{max}(t^*) = q_2$. By Lemma 33, the pair $(k_r, k_t) \in [0,1]\times[0,1]$ exists and is unique. By Lemma 32, the associated pair (r^*, t^*) is also unique.

By Lemmas 27 and 30, there exists a sequence of Ellsberg strategy profiles $e(\epsilon) = \big([p_{min}(\epsilon), p_{max}(\epsilon)], [q_{min}(\epsilon), q_{max}(\epsilon)]\big)$ induced by equilibria in the sequence of disturbed games such that

$$\lim_{\epsilon\to 0} p_{max}(\epsilon) = p^*, \textit{and}$$
$$\lim_{\epsilon\to 0} q_{max}(\epsilon) = q^*.$$

By Lemma 32, this can only be the case if the sequence of thresholds $\{(r^*(\epsilon), t^*(\epsilon))\}$ associated to the sequence of maxmin strategies equilibria by conditions (4.15) to (4.18) is such that

$$\lim_{\epsilon\to 0}\{(r^*(\epsilon), t^*(\epsilon))\} = (r^*, t^*).$$

By the construction of (k_r, k_t), this implies

$$\lim_{\epsilon\to 0} p_{min}(\epsilon) = p_1, \textit{and}$$
$$\lim_{\epsilon\to 0} q_{min}(\epsilon) = q_1.$$

This shows $\lim_{\epsilon\to 0} e(\epsilon) = e$ and the equilibrium is disambiguable.

■

4.3.6 Adapting the proof for other games in $\mathbf{\Gamma}$

The proof presented was designed for games in $\mathbf{\Gamma}^{D-4}$. It did not cover games of class I nor some games in class II. We do not provide here a proof for those games but discuss shortly what parts of the proof need to be adatped.

Extending the proof to games in $\mathbf{\Gamma}^{II-D}$ that do not belong to $\mathbf{\Gamma}^{D-4}$ is easy. The unique adaptations relates to the extreme point occupied by the indifference strategy in the equilibrium Ellsberg strategy. These extreme points are given in Table 4.1.

Γ_3		q L	1-q R
p	U	$0,0$	$2,1$
1-p	D	$1,2$	$0,0$

Figure 4.4: Symmetric "Battle of the Sexes" game $\Gamma_3 \in \mathbf{\Gamma}^{D-4}$ for which $p^* = \frac{2}{3}$, $\bar{p} = \frac{1}{3}$, $q^* = \frac{2}{3}$ and $\bar{q} = \frac{1}{3}$.

Games of class I are such that their disturbed versions admit equilibria in pure strategies. The conditions under which a pure strategy in the disturbed game is a best reply are given in Appendix 4.6.4. These conditions rely on a unique threshold and only pure strategies of the basic games are used. As a result, equilibrium conditions are simpler than expressed in Lemma 26. The only difficulty is to select the appropriate threshold among the two thresholds implied by the induced Ellsberg strategies. The selection procedure is given in Lemma 34 (see Appendix 4.6.4). There is no need to search for simpler conditions for small disturbances as only one threshold defines a pure strategy in the disturbed game. The proof of existence of equilibrium follows exactly the same lines. The major difference is that A and B are replaced respectively by 0 and 1 (and so are C and D).

Games of class II that do not belong to $\mathbf{\Gamma}^{II-D}$ are hybrid in the sense that the best reply of one player is a pure strategy whereas the one of the other is a maxmin strategy. As a result, the proof for this case will borrow elements from the proof for games of class I and games in $\mathbf{\Gamma}^{II-D}$.

4.4 An example of disambiguation

In this section, we illustrate the disambiguation of proper Ellsberg equilibria in the symmetric game $\Gamma_3 \in \mathbf{\Gamma}^{D-4}$ of the type "Battle of the sexes", illustrated in Figure 4.4. For this game, the expected utilities are given by:

$$U_1(p,q) = 2p + q(1-3p),$$
$$U_2(q,p) = 2q + p(1-3q).$$

The indifference strategies are $p^* = q^* = \frac{2}{3}$ and the maxmin strategies are $\bar{p} = \bar{q} = \frac{1}{3}$, which confirms that Γ_3 belongs to $\mathbf{\Gamma}^{D-4}$. Riedel and Sass (2013) have shown that for this coordination game, the set of Ellsberg equilibrium is $\left\{\left([p_1, \frac{2}{3}], [q_1, \frac{2}{3}]\right) \,\middle|\, \frac{1}{3} \leq p_1, q_1\right\}$.[10]

[10] Our game Γ_3 belongs to the set of games Riedel and Sass (2013) cover if one inverses the two columns. Therefore, Γ_3 is such that $\bar{p} \leq p^*$ and $(1-q)^* \leq \overline{(1-q)}$.

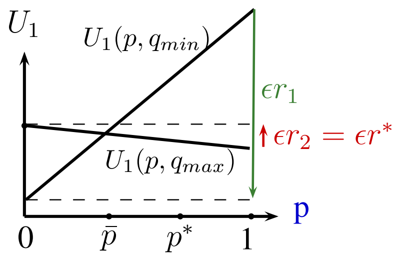

Figure 4.5: Thresholds value r_1 and r_2 for a given Ellsberg strategy $[q_{min}, q_{max}]$.

4.4.1 Equilibria in a disturbed game

Consider the disturbed game $\Gamma^*(\epsilon)$ obtained by attaching disturbances ϵr and ϵt to the payoffs associated to the pure strategies U and L respectively. The realization of the disturbances is private information. The domains $\mathcal{P}_r$ and $\mathcal{P}_t$, in which the probability distributions of the ambiguous variables r and t lie, are common knowledge. Let their basic densities f_r^b and f_t^b be the uniform densities on $[-1, 1]$:

$$\mathcal{P}_r := \left\{ f \in \mathcal{F} \,\middle|\, \frac{1-k_r}{2} \leq f(x) \leq \frac{1+k_r}{2} \text{ for all } x \in [-1,1] \right\},$$
$$\mathcal{P}_t := \left\{ f \in \mathcal{F} \,\middle|\, \frac{1-k_t}{2} \leq f(x) \leq \frac{1+k_t}{2} \text{ for all } x \in [-1,1] \right\}.$$

We compute equilibrium strategies in the disturbed game, as a function of the ambiguity parameters k_r and k_t. As shown in Lemma 24, the strategy picked by player 2 will appear to player 1 as an Ellsberg strategy $[q_{min}, q_{max}]$. By Lemma 25, her best reply to $[q_{min}, q_{max}]$ is a maxmin strategy characterized by two thresholds r_1 and r_2. These two thresholds are illustrated in Figure 4.5. As shown in Lemma 30, for sufficiently small ϵ, a unique threshold value lies in the support of r. This threshold r^* is the one associated with q_{max}, because $q^* = \frac{2}{3}$ is the upper bound of the equilibrium Ellsberg strategy of player 2. The best reply to $[q_{min}, q_{max}]$ is a pure strategy whose threshold r^* is obtained in equation (4.27). For player 2, the best reply to $[p_{min}, p_{max}]$ is a pure strategy whose threshold t^* is obtained in equation (4.28).

$$\epsilon r^* = U_1(0, q_{max}) - U_1(1, q_{max}) = 3q_{max} - 2, \tag{4.27}$$
$$\epsilon t^* = U_2(0, p_{max}) - U_2(1, p_{max}) = 3p_{max} - 2. \tag{4.28}$$

Thresholds r^* and t^* completely characterize the equilibrium strategies of the players for small disturbances. Our objective is therefore to compute their values as a function of the parameters of the game.

By simplifying equation (4.14) in Lemma 26, we have:[11]

$$\begin{aligned} q_{max} &= \max_{f \in \mathcal{P}_t} \int_{-1}^{t^*} \frac{1}{3} f(t)dt + \int_{t^*}^{1} 1 f(t)dt \\ &= \left(1 - (1-t^*)\frac{1+k_t}{2}\right)\frac{1}{3} + (1-t^*)\frac{1+k_t}{2}1 \\ &= \frac{1}{3}\big(2 - t^*(1+k_t) + k_t\big) \end{aligned}$$

Similarly, we obtain for p_{max}:

$$p_{max} = \max_{f \in \mathcal{P}_r} \int_{-1}^{r^*} \frac{1}{3} f(r)dr + \int_{r^*}^{1} 1 f(r)dr = \frac{1}{3}\big(2 - r^*(1+k_r) + k_r\big)$$

Replacing in equations (4.27) and (4.28), the values of q_{max} and p_{max} found above, we obtain a system of two equations:

$$\begin{cases} \epsilon r^* &= k_t - t^*(1+k_t), \\ \epsilon t^* &= k_r - r^*(1+k_r). \end{cases}$$

Solving this system yields:

$$\begin{cases} r^* &= \frac{\epsilon k_t - k_r(1+k_t)}{\epsilon^2 - (1+k_r)(1+k_t)}, \\ t^* &= \frac{\epsilon k_r - k_t(1+k_r)}{\epsilon^2 - (1+k_t)(1+k_r)}. \end{cases}$$

This last system of equations characterizes for small ϵ the equilibrium strategies in the disturbed game, as a function of the size of the disturbance ϵ and the ambiguity parameters k_r and k_t. When the disturbance ϵ tends to 0, we have:

$$\lim_{\epsilon \to 0} r^* = \frac{k_r}{1+k_r}, \tag{4.29}$$

$$\lim_{\epsilon \to 0} t^* = \frac{k_t}{1+k_t}. \tag{4.30}$$

4.4.2 The sequence of equilibria

The equilibrium strategy of player 2, characterized by t^*, is perceived by player 1 as an Ellsberg strategy $[q_{min}, q_{max}]$. Alternatively, player 2 perceives the strategy of player 1 characterized by r^* as an Ellsberg strategy $[p_{min}, p_{max}]$. We computed above that:

$$q_{max} = \frac{1}{3}\big(2 - t^*(1+k_t) + k_t\big), \tag{4.31}$$

$$p_{max} = \frac{1}{3}\big(2 - r^*(1+k_r) + k_r\big). \tag{4.32}$$

[11] The second expression is valid when t^* has a positive value, which is verified for small ϵ by equation (4.30).

Similarly, we derive:

$$q_{min} = \frac{1}{3}\big(2 - t^*(1-k_t) - k_t\big), \tag{4.33}$$

$$p_{min} = \frac{1}{3}\big(2 - r^*(1-k_r) - k_r\big). \tag{4.34}$$

Replacing in equations (4.31) to (4.34) variables t^* and r^* by their values allows to compute the sequence $\{e(\epsilon)\}$ of induced Ellsberg strategy profiles in the disturbed game $\Gamma^*(\epsilon)$. Theorem 11 proved that for all proper equilibria e of Γ, we can find a pair $(k_r, k_t) \in [0,1]\times[0,1]$ such that $\lim_{\epsilon\to 0}\{e(\epsilon)\} = e$. Replacing r^* and t^* by their value in the limit, we obtain:

$$\lim_{\epsilon\to 0}\big([p_{min}, p_{max}], [q_{min}, q_{max}]\big) = \left(\left[\frac{2}{3}\frac{1}{1+k_r}, \frac{2}{3}\right], \left[\frac{2}{3}\frac{1}{1+k_t}, \frac{2}{3}\right]\right) \tag{4.35}$$

Remember that all proper Ellsberg equilibrium of this game are of the form $e = \big([p_1, \frac{2}{3}], [q_1, \frac{2}{3}]\big)$ with $\frac{1}{3} \leq p_1, q_1$. For $k_r = 0$, we have $p_{min} = \frac{2}{3}$ and for $k_r = 1$, we have $p_{min} = \frac{1}{3}$. Since function p_{min} is strictly monotone in k_r between these two bounds, and since the same is true for function q_{min} and k_t, any Ellsberg equilibrium in Γ_3 can be disambiguated for a unique pair (k_r, k_t).

4.5 Concluding Remarks

Riedel and Sass (2013) have introduced Ellsberg games and proposed a solution concept that they call Ellsberg equilibrium. It is a coarsening of Nash equilibrium. Any Nash equilibrium is an Ellsberg equilibrium but the converse does not hold: (quasi-) proper Ellsberg equilibria are not Nash equilibria. For the class of 2×2 normal form games that we consider, Harsanyi (1973) has shown that all Nash equilibria in mixed strategies can be purified. Our Disambiguation Theorem shows that all (quasi-) proper Ellsberg equilibria can be disambiguated. Moreover, for games of class I, all (quasi-) proper Ellsberg equilibria can be purified. In that sense, our result extends that of Harsanyi.

Generalizing our theorem beyond 2×2 normal games can unfortunately not be done using the mathematical technique of Harsanyi. In effect, ambiguity averse players perform non-smooth evaluations of ambiguous outcomes. We can nevertheless see no fundamental reason why this generalization could not be performed, even though some challenging obstacles need to be overcome.

4.6 Appendix

4.6.1 Proof of Lemma 22

Proof. A proof of claim 1 can be found in Fudenberg and Maskin (1991). We prove claim 2: if player 1's payoffs are not Column Dominant in game Γ then $\bar{p} \in (0, 1)$

and $U_1(\bar{p}, q) = U_1(\bar{p}, q')$ for all $q, q' \in [0,1]$.

Geometrically, given the strategy q chosen by player 2, the expected utility $U_1^q(p) := U_1(p,q)$ defines a line in $[0,1] \times \mathbb{R}$. If we allow the domain of q to be $\mathbb{R}$, this line is defined in $\mathbb{R}^2$. The family of such lines $\{U_1^q(p)\}_{q \in [0,1]}$ has the property of Unique Intersection.

Property 1 (Unique Intersection)**.**
Let $\{U_1^q(p)\}_{q\in[0,1]}$ be a family of lines defined in $\mathbb{R}^2$. The family has the property of unique intersection if there exists a point $(\tilde{p}, u_1) \in \mathbb{R}^2$ at which all members of the family intersect.

This unique intersection $(\tilde{p}, u_1)$ is hence such that for all $q \in [0,1]$, the point $(\tilde{p}, u_1) \in U_1^q(p)$.

We show now that family $\{U_1^q(p)\}_{q\in[0,1]}$ has a unique intersection $(\tilde{p}, u_1)$. Player 1's expected utility can be rewritten:

$$U_1(p,q) = \pi_1^4 + q\big((\pi_1^3 - \pi_1^4) + p(\pi_1^1 - \pi_1^2 - \pi_1^3 + \pi_1^4)\big) + p(\pi_1^2 - \pi_1^4).$$

The value $\tilde{p}$ at which an intersection takes place is therefore the solution of the following equation:

$$(\pi_1^3 - \pi_1^4) + \tilde{p}(\pi_1^1 - \pi_1^2 - \pi_1^3 + \pi_1^4) = 0.$$

As there are no weakly dominant strategies in Γ, player 1's payoff are not Row Dominant and hence two cases can arise.

- Case A: $\pi_1^1 > \pi_1^3$ and $\pi_1^2 < \pi_1^4$.

  The solution $\tilde{p}$ of last equation belongs to $(0,1)$ if either $\pi_1^1 > \pi_1^2$ and $\pi_1^3 < \pi_1^4$ or $\pi_1^1 < \pi_1^2$ and $\pi_1^3 > \pi_1^4$. This means $\tilde{p} \in (0,1)$ if player 1's payoff are not Column Dominant. Therefore the factor $\pi_1^1 - \pi_1^2 - \pi_1^3 + \pi_1^4$ is different from zero. As a result the solution $\tilde{p}$ is unique.

- Case B: $\pi_1^1 < \pi_1^3$ and $\pi_1^2 > \pi_1^4$

  A parallel argument can be made to show $\tilde{p} \in (0,1)$ and is unique.

We now prove for games with $\tilde{p} \in (0,1)$ that this intersection is the maxmin strategy, that is $\tilde{p} = \bar{p}$. By definition of indifference strategy q^*, we have for all $p \in [0,1]$ that $U_1(p,q^*) = U_1(\tilde{p}, q^*) = u_1$ and hence $U_1^{q^*}(p)$ is flat: $U_1(0,q^*) - U_1(1,q^*) = 0$. As player 1 has no weakly dominant strategy, the difference $U_1(0,q) - U_1(1,q)$ is strictly monotone in q. This implies q^* is the only value for which $U_1^q(p)$ is flat. We showed that $q^* \in (0,1)$, implying there exist hence q' and q'' in $[0,1]$ such that $q' < q^* < q''$. By strict monotonicity of the difference $U_1(0,q) - U_1(1,q)$, we have that among the two lines $U_1^{q'}(p)$ and $U_1^{q''}(p)$, one is strictly increasing and the other strictly decreasing. Therefore, as $\tilde{p} \in (0,1)$, for any $p \in [0,1]$ with $p \neq \tilde{p}$ there exists $q \in [0,1]$ with $q \neq q^*$ such that $U_1^q(p) < U_1^q(\tilde{p})$. The maxmin strategy $\bar{p}$ is hence at the intersection $\tilde{p}$. This completes the proof as we showed that utility in $\tilde{p}$ is independent of q. The proof of claim 3 is done using the same argument.

■

4.6.2 Proof of Lemma 23

Proof. We assume without loss of generality that $p_1 < p_2$. First, we show $q^* \in \{q_1, q_2\}$. As shown in Lemma 22, for all $\Gamma \in \mathbf{\Gamma}$, we have that $q^* \in (0,1)$ and is unique. The Ellsberg strategy $[p_1, p_2]$ is a best reply to $[q_1, q_2]$, if and only if we have for all $p \in [p_1, p_2]$ there exists no $p' \in [0,1]$ such that $U_1\big(p', [q_1, q_2]\big) > U_1\big(p, [q_1, q_2]\big)$. Being ambiguity averse, player 1 must be indifferent between all mixed strategies inside the Ellsberg strategy $[p_1, p_2]$ she plays. Formally, for all $p, p' \in [p_1, p_2]$ we have

$$U_1\big(p', [q_1, q_2]\big) = U_1\big(p, [q_1, q_2]\big).$$

As $U_1\big(p, [q_1, q_2]\big) = \min\big(U_1(p, q_1), U_1(p, q_2)\big)$ (equation (4.1)), we must have either

- $U_1(p, q_1)$ is constant (implying $q_1 = q^*$) and $U_1(p, q_1) \leq U_1(p, q_2)$ for all $p \in [p_1, p_2]$, or
- $U_1(p, q_2)$ is constant (implying $q_2 = q^*$) and $U_1(p, q_2) \leq U_1(p, q_1)$ for all $p \in [p_1, p_2]$.

Therefore we have $q^* \in \{q_1, q_2\}$.

Second, we show $p^* \in \{p_1, p_2\}$. If $[q_1, q_2]$ is a proper Ellsberg strategy, then the reasoning above proves it. We show it holds as well if the equilibrium is quasi-proper, that is $q_1 = q_2$. From the previous reasoning, this implies $q_1 = q_2 = q^*$. The Ellsberg strategy q^* is a best reply to $[p_1, p_2]$, if and only if there exists no $q' \in [0,1]$ such that $U_2\big(q', [p_1, p_2]\big) > U_2\big(q^*, [p_1, p_2]\big)$. Remember we have $U_2\big(q, [p_1, p_2]\big) = \min\big(U_2(q, p_1), U_2(q, p_2)\big)$. Function $U_2(q, p)$ is linear in q. We show that if $p^* \notin \{p_1, p_2\}$, we have a contradiction.

- If $U_2(q, [p_1, p_2])$ is strictly increasing in q on $[0,1]$, then best reply is $q_1 = q_2 = 1$, and since $q^* \neq 1$ for all $\Gamma \in \mathbf{\Gamma}$, we have a contradiction.
- If $U_2(q, [p_1, p_2])$ is strictly decreasing in q on $[0,1]$, then best reply is $q_1 = q_2 = 0$, and since $q^* \neq 0$, we have another contradiction.
- If $U_2(q, [p_1, p_2])$ is strictly increasing in q on one portion of $[0,1]$ and stricly decreasing on the other, then the best reply is $\bar{q}$. In effect, by the property of Unique Intersection, $U_2(q, p_1)$ and $U_2(q, p_2)$ must then intersect in $(\bar{q}, u_2)$ and $U_2(q, [p_1, p_2])$ has maximal value for $q = \bar{q}$. As for all $\Gamma \in \mathbf{\Gamma}$, $q^* \neq \bar{q}$, we have yet another contradiction.

The only possibility for q^* to belong to best replies is that either $U_2(q, p_1)$ or $U_2(q, p_2)$ is constant in q, which implies $p^* \in \{p_1, p_2\}$. ∎

4.6.3 Proof of Lemma 24

Proof. We show that for all $p \in [p_{min}, p_{max}]$, there exists a density $f \in \mathcal{P}_r$ such that

$$p = \int_{-1}^{1} p^b(r) f(r) dr.$$

The domain $\mathcal{P}_r$ is convex. This means that for all $f_1, f_2 \in \mathcal{P}_r$, distribution f_3 defined as $f_3(r) := \lambda f_1(r) + (1-\lambda) f_2(r)$ belongs to $\mathcal{P}_r$. The mapping $\int_{-1}^{1} p^b(r) f(r) dr$ is linear in f. As the image of a convex set through a linear mapping is a convex set, the image of $\mathcal{P}_r$ through this mapping is convex. In the real line, a convex set is an interval. As $\mathcal{P}_r$ is closed, so must be its image $[p_{min}, p_{max}]$. ■

4.6.4 Best replies in disturbed game

For all $\Gamma \in \mathbf{\Gamma}$ such that player 1's payoffs are Column Dominant, the following lemma provides conditions under which a pure strategy of player 1 is a best reply to a strategy of player 2 inducing $[q_{min}, q_{max}]$.[12]

Lemma 34 (Best-Reply in pure strategy)**.**
For all $\epsilon > 0$ and all $\Gamma \in \mathbf{\Gamma}$ such that player 1's payoffs are Column Dominant, the strategy p^b is a best reply to any $[q_{min}, q_{max}] \subseteq [0,1]$ if it is a pure strategy $p^b = r^* \in S^1_{pu}$ defined by:

$$\begin{aligned} \epsilon r' &= U_1(0, q_{min}) - U_1(1, q_{min}), \\ \epsilon r'' &= U_1(0, q_{max}) - U_1(1, q_{max}), \\ r^* &= \begin{cases} max(r', r'') & \textit{if } |\pi_1^3 - \pi_1^4| \leq |\pi_1^1 - \pi_1^2|, \\ min(r', r'') & \textit{otherwise.} \end{cases} \end{aligned}$$

Proof. The proof is only provided for the conditions on player 1's payoffs leading to $r^* = max(r', r'')$. Those conditions ensure that the solution to equation $(\pi_1^3 - \pi_1^4) + p(\pi_1^1 - \pi_1^2 - \pi_1^3 + \pi_1^4) = 0$, which yields the unique intersection $\tilde{p}$, is non-positive. Therefore the relevant threshold among r' and r'' is the largest one.[13]

Take any $\epsilon > 0$, any $[q_{min}, q_{max}] \subseteq [0,1]$ and any $\Gamma \in \mathbf{\Gamma}$ such that

- player 1's payoffs are Column Dominant, and
- the solution to equation $(\pi_1^3 - \pi_1^4) + p(\pi_1^1 - \pi_1^2 - \pi_1^3 + \pi_1^4) = 0$ is non-positive.

Given equations (4.1) and (4.2), we have

$$\begin{aligned} U_1\big(p, [q_{min}, q_{max}], \epsilon r\big) &= min\big(U_1(p, q_{min}, \epsilon r), U_1(p, q_{max}, \epsilon r)\big) \\ &= min\big(U_1(p, q_{min}) + p\epsilon r, U_1(p, q_{max}) + p\epsilon r\big) \end{aligned}$$

where $U_1(p, q, \epsilon r)$ is linear in p since $U_1(p, q)$ is linear in p. For the considered Γ, $\bar{p}$ is not a proper mixed strategy and the unique intersection of $U_1(p, q_1)$ and $U_1(p, q_2)$ is in $(\tilde{p}, u_1)$ with $\tilde{p} \leq 0$. Therefore, if $q_{min} \neq q_{max}$, we have two possible cases:

[12] In Lemma 34, the conditions under which the expression $max(r', r'')$ is used for r^* are such that the solution p of the implicit equation $U_1(p, 0) = U_1(p, 1)$ is non-positive. The conditions under which the expression $min(r', r'')$ is used for r^* are such that the solution p of the implicit equation $U_1(p, 0) = U_1(p, 1)$ is equal or larger than 1.

[13] If $q_{min} = q_{max}$ then $r^* = r' = r''$. But if $q_{min} < q_{max}$, two cases can arise: either $U_1(p, q_{min}) < U_1(p, q_{max})$ for all $p \in (0, 1]$ or $U_1(p, q_{min}) > U_1(p, q_{max})$ for all $p \in (0, 1]$. Assume that the game Γ is such that the first of these two cases arises. The relevant threshold r^* is therefore r' associated to q_{min}. Under the payoff conditions leading to $r^* = max(r', r'')$, $U_1(p, q_{min})$ and $U_1(p, q_{max})$ are two straight lines which cross at $\tilde{p} \leq 0$. As a result, we have $U_1(0, q_{min}) - U_1(1, q_{min}) > U_1(0, q_{max}) - U_1(1, q_{max})$. The other case leads to the same conclusion.

- Case A: $U_1(p, q_{min}) > U_1(p, q_{max})$ for all $p \in (0, 1)$,
- Case B: $U_1(p, q_{min}) < U_1(p, q_{max})$ for all $p \in (0, 1)$.

Let $\hat{q} := q_{max}$ if we are in case A and $\hat{q} := q_{min}$ if we are in case B. The mixed strategy $\hat{q} \in \{q_{min}, q_{max}\}$ is the one associated with the minimal utility for player 1, the one she takes into account in front of ambiguity. We have hence

$$U_1\big(p, [q_{min}, q_{max}], \epsilon r\big) = U_1(p, \hat{q}) + p\epsilon r.$$

From the definition of $\hat{q}$ and the definition of r^* in the statement of the lemma, we have $\epsilon r^* = U_1(0, \hat{q}) - U_1(1, \hat{q})$, which can be rewritten $U_1(0, \hat{q}) + 0\epsilon r^* = U_1(1, \hat{q}) + 1\epsilon r^*$ implying that:

1. $U_1(0, \hat{q}) + 0\epsilon r > U_1(1, \hat{q}) + 1\epsilon r$ for all $r < r^*$,
2. $U_1(0, \hat{q}) + 0\epsilon r < U_1(1, \hat{q}) + 1\epsilon r$ for all $r > r^*$.

As $U_1(p, \hat{q}) + p\epsilon r$ is a linear function of p, the best reply to all $r < r^*$ is $p = 0$ and the best reply to all $r > r^*$ is $p = 1$. If $r = r^*$ then $U_1(p, \hat{q}) + p\epsilon r$ is a constant and any $p \in [0, 1]$ is a best reply, and in particular $p = 0$. ∎

For all $\Gamma \in \mathbf{\Gamma}$ such that player 2's payoffs are Row Dominant, there exists similar conditions under which a pure strategy of player 2 is a best reply to a strategy of player 1 inducing $[p_{min}, p_{max}]$.[14]

[14] For for all $\Gamma \in \mathbf{\Gamma}$ such that player 2's payoffs are Row Dominant, those conditions are obtained for pure strategies of player 2 by replacing π_1^1 by π_2^1, π_1^4 by π_2^4, π_1^2 by π_2^3 and π_1^3 by π_2^2.

Bibliography

Aleskerov, F. and Kurbanov, E. (1999). Degree of manipulability of social choice procedures. In Alkan, P. A., Aliprantis, P. C. D., and Yannelis, P. N. C., editors, *Current Trends in Economics*, number 8 in Studies in Economic Theory, pages 13–27. Springer Berlin Heidelberg.

Anderson, E. and Esposito, L. (2013). On the joint evaluation of absolute and relative deprivation. *The Journal of Economic Inequality*, 12(3):411–428.

Andersson, T., Ehlers, L., and Svensson, L.-G. (2014). Least manipulable Envy-free rules in economies with indivisibilities. *Mathematical Social Sciences*, 69:43–49.

Araar, A. and Duclos, J. Y. (2009). Testing For Pro-Poorness of Growth, with an Application to Mexico. *Review of Income and Wealth*, 55(4):853–881.

Archibald, G. and Donaldson, D. (1976). Paternalism and prices. In Allingham, M. and Burstein, M., editors, *Resource allocation and economic policy*. London: Macmillan.

Arribillaga, R. P., Massó, J., and others (2014). Comparing generalized median voter schemes according to their manipulability. Technical report.

Arrow, K. (1950). A difficulty in the concept of social welfare. *The Journal of Political Economy*, 58(4):328–346.

Atkinson, A. (1970). On the Measurement of Inequality. *Journal of Economic Theory*, 2:244–263.

Atkinson, A. and Bourguignon, F. (2001). Poverty and Inclusion from a World Perspective. In Stiglitz, J. and Muet, P.-A., editors, *Governance, equity and global markets*. Oxford University Press, New York.

Barberà, S. (2014). Choosing with an Agenda. (2012):1–39.

Barberà, S. and Gerber, A. (2014). Sequential Voting and Agenda Manipulation. SSRN Scholarly Paper ID 2483329, Social Science Research Network, Rochester, NY.

Blank, R. M. (2008). Presidential address : How to Improve Poverty United States. 27(2):233–254.

Bolton, G. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193.

Bourguignon, F. (2013). *La Mondialisation de l'Inégalité*.

Brennan, G. (1973). Pareto desirable redistribution: the case of malice and envy. *Journal of Public Economics*, 2:173–183.

Camerer, C. and Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5(4):325–370.

Chakravarty, S. R. (1983). A New Index of Poverty. *Mathematical Social Science*, 6:307–313.

Charness, G. and Rabin, M. (2002). Understanding Social Preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chen, D. and Schonger, M. (2012). Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations. (January).

Chen, S. and Ravallion, M. (2013). More Relatively-Poor People in a Less Absolutely-Poor World. *Review of Income and Wealth*, 59(1):1–28.

Citro, C. F. and Michael, R. T. (1995). Measuring Poverty: A new approach.

Corazzini, L., Esposito, L., and Majorano, F. (2011). Exploring the Absolutist vs Relativist Perception of Poverty using a Cross-Country Questionnaire Survey. *Journal of Economic Psychology*, 32(2):273–283.

Cowell, F. A. and Victoria-Feser, M. P. (1994). Robustness properties of poverty indices.

Dasgupta, P. and Maskin, E. (2008). On the robustness of majority rule. *Journal of the European Economic Association*, 6(5):949–973.

Dasgupta, P. and Maskin, E. (2010). Elections and Strategic Voting: Condorcet and Borda. *Unpublished slides, UC Irvine Conference on Adaptive Systems and Mechanism Design.*

Day, R. and Milgrom, P. (2008). Core-selecting package auctions. *international Journal of game Theory*, 36(3-4):393–407.

de Mesnard, L. (2007). Poverty Reduction: The Paradox of the Endogenous Poverty Line.

Decerf, B. and Van der Linden, M. (2014). Fair social orderings with other-regarding preferences. *Mimeo.*

Duclos, J. Y. (2003). What is Pro-Poor? *Social Choice and Welfare*, 32(1):37–58.

Duclos, J. Y. and Gregoire, P. (2002). Absolute and relative deprivation and the measurement of poverty. *Review of Income and Wealth*, 48(4):471–492.

Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., and Sobel, J. (2011). Other-Regarding Preferences in General Equilibrium. *The Review of Economic Studies*, 78(2):613–639.

Easton, B. (2002). Beware the Median. *Social Policy Research Center Newsletter*, 82:6–7.

Edgeworth, F. Y. (1881). *Mathematical Psychics*. Augustus M. Kelley.

Eichberger, J. and Kelsey, D. (2000). Non-Additive Beliefs and Strategic Equilibria. *Games and Economic Behavior*, 30(2):183–215.

Erdil, A. and Klemperer, P. (2010). A new Payment Rule for Core-selecting Package Auctions. *Journal of the European Economic Association*, 8(2-3):537–547.

Ergin, H. and Sönmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics*, 90(1-2):215–237.

European Commission (2015). *Portfolio of Indicators for the Monitoring of the European Strategy for Social Protection and Social Inclusion*. European Commission, Brussels.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114:817–868.

Fleurbaey, M. (2012). The importance of what people care about. *Politics, Philosophy & Economics*, 11(4):415–447.

Fleurbaey, M. and Maniquet, F. (2006). Fair social orderings. *Economic Theory*, 34(1):25–45.

Fleurbaey, M. and Maniquet, F. (2011). *A Theory of Fairness and Social Welfare*. Cambridge University Press.

Fleurbaey, M. and Schokkaert, E. (2011). Behavioral fair social choice. *Core discussion paper*.

Fleurbaey, M. and Trannoy, A. (2003). The impossibility of a Paretian egalitarian. *Social Choice and Welfare*, 21(2):243–263.

Foster, J., Greer, J., and Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica*, 52(3):761–766.

Foster, J., Seth, S., Lokshin, M., and Sajaia, Z. (2013). *A Unified Approach to Measuring Poverty and Inequality: Theory and Practice*. World Bank, Washington, DC.

Foster, J. and Shorrocks, A. (1991). Subgroup Consistent Poverty Indices. *Econometrica*, 59(3):687–709.

Foster, J. E. (1998). Absolute versus Relative Poverty. *American Economic Review*, 88(2):335–341.

Foster, J. E. and Sen, A. (1997). On Economic Inequality after a Quarter Century. In Sen, A., editor, *On Economic Inequality*. Clarendon Press, Oxford, rev. ed. edition.

Foster, J. E. and Szekely, M. (2008). Is Economic Growth Good for the Poor? Tracking Low Incomes Using General Means. *International Economic Review*, 49(4):1143–1172.

Fudenberg, D. and Maskin, E. (1991). On the dispensability of public randomization in discounted repeated games. *Journal of Economic Theory*, 428438:428–438.

Fujinaka, Y. and Wakayama, T. (2015). Maximal manipulation of envy-free solutions in economies with indivisible goods and money. *Journal of Economic Theory*, 158, Part A:165–185.

Gaertner, W. and Schokkaert, E. (2012). *Empirical Social Choice: Questionnaire: Experimental Studies on Distributive Justice*. Number November. Cambridge University Press.

Gajdos, T., Hayashi, T., Tallon, J.-M., and Vergnaud, J.-C. (2008). Attitude toward imprecise information. *Journal of Economic Theory*, 140(1):27–65.

Gale, D. and Shapley, L. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Gersbach, H. and Haller, H. (2001). Collective Decisions and Competitive Markets. *The Review of Economic Studies*, 68(2):347–368.

Gilboa, I. and Schmeidler, D. (1989). Maximin Expected Utility with non-Unique Priors. *Journal of Mathematical Economics*, 18(2):141 – 153.

Goodin, R. E. (1986). Laundering preferences. In Elster, J. and Hyland, A., editors, *Foundations of Social Choice Theory*, pages 75–102. Cambridge University Press.

Gorman, W. M. (1968). Symposium on Aggregation: The Structure of Utility Functions. *Review of Economic Studies*, 35(4):367–390.

Greenberg, J. (2000). The right to remain silent. *Theory and Decision*, 48(2):193–204.

Haeringer, G. and Klijn, F. (2008). Constrained School Choice. Technical Report 294.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

Hammond, P. J. (1987). altruism. In Eatwell, J., Milgate, M., and Newman, P., editors, *The New Palgrave: A Dictionary of Economics. First Edition*, pages 1–8. Palgrave Macmillan.

Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23.

Harsanyi, J. C. (1982). Morality and the theory of rationale behaviour. In Sen, A. and William, B., editors, *Utilitarianism and beyond*, pages 39–62. Cambridge University Press.

Hochman, H. and Rodgers, J. (1969). Pareto Optimal Redistribution. *The American economic review*, 59:542–557.

Hochman, H. and Rodgers, J. (1971). Is efficiency a criterion for judging redistribution? *Public Finance*, 26:348– 360.

Huber, P., J. (1981). *Robust Statistics*. John Wiley, New-York.

Kakwani, N. C. (1980). On a Class of Poverty Measures. *Econometrica*, 48(2):437–446.

Kakwani, N. C. (2008). Poverty Equivalent Growth Rate. *Review of Income and Wealth*, 54(4):643–655.

Kesten, O. (2010). School choice with consent. *The Quarterly Journal of Economics*, (August):1297–1348.

Kesten, O. (2011). On two kinds of manipulation for school choice problems. *Economic Theory*, 51(3):677–693.

Kojima, F. and Manea, M. (2010). Axioms for Deferred Acceptance. *Econometrica*, 78(2):633–653.

Konow, J. (2001). Fair and square: the four sides of distributive justice. *Journal of Economic Behavior & Organization*, 46(2):137–164.

Lo, K. C. (1996). Equilibrium in Beliefs under Uncertainty. *Journal of Economic Theory*, 71(2):443–484.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498.

Madden, D. (2000). Relative or absolute poverty lines: a new approach. *Review of Income and Wealth*, 46(2):181–199.

Marinacci, M. (2000). Ambiguous Games. *Games and Economic Behavior*, 31(2):191–219.

Mas-Collel, A., Green, J. R., and Whinston, M. D. (1995). *Microeconomic Theory*. Oxford University Press.

Maus, S., Peters, H., and Storcken, T. (2007a). Anonymous voting and minimal manipulability. *Journal of Economic Theory*, 135(1):533–544.

Maus, S., Peters, H., and Storcken, T. (2007b). Minimal manipulability: anonymity and unanimity. *Social Choice and Welfare*, 29(2):247–269.

Neumann, J. V. (1928). Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320.

Parkes, D. C., Kalagnanam, J., and Eso, M. (2002). Achieving budget-balance with VCG-based payment schemes in combinatorial exchanges. Technical report, Technical report, IBM Research RC22218 W0110-065.

Pathak, P. A. and Sönmez, T. (2013). School Admissions Reform in Chicago and England : Comparing Mechanisms by their Vulnerability to Manipulation. *American Economic Review*, 103(1):80–106.

Pazner, E. and Schmeidler, D. (1978). Egalitarian equivalent allocations: A new concept of economic equity. *Quarterly Journal of Economics*, 92:671–687.

Ravallion, M. (2008). On the Welfarist Rationale for Relative Poverty Lines. *Policy Research Working Paper 484*, 4486(January).

Ravallion, M. (2012). Poverty Lines Across the World. In Jefferson, P. N., editor, *Oxford Handbook of the Economics of Poverty*, number January. Oxford University Press, Oxford.

Ravallion, M. and Chen, S. (2003). Measuring Pro-poor Growth. *Economics Letters*, 78(1):93–99.

Ravallion, M. and Chen, S. (2011). Weakly relative poverty. *Review of Economics and Statistics*, 93(4):1251–1261.

Ravallion, M., Chen, S., and Sangraula, P. (2009). Dollar a Day Revisited. *The World Bank Economic Review*, 23(2):163–184.

Riedel, F. and Sass, L. (2013). Ellsberg games. *Theory and Decision*, 76(4):469–509.

Roth, A. E. and Rothblum, U. G. (1999). Truncation Strategies in Matching Markets-in Search of Advice for Participants. *Econometrica*, 67(1):21–43.

Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill.

Ruggles, P. (1990). Drawing the line.

Runciman, W. G. (1966). *Relative Deprivation and Social Justice*. Routledge, London.

Schelling, T. C. (1980). *The strategy of conflict*. Harvard university press.

Schokkaert, E. (1999). M . Tout-le-monde est "post-welfariste ". Opinions sur la justice redistributive. *Revue économique*, 50:811–831.

Sen, A. (1976). poverty an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*.

Sen, A. (1983). Poor , Relatively Speaking. *Oxford Economic Papers*, 35(2):153–169.

Sen, A. (1992). *Inequality Reexamined*. Harvard University Press, Cambridge.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Home University Library, everydayman edition edition.

Sprumont, Y. (2012). Resource egalitarianism with a dash of efficiency. *Journal of Economic Theory*, 147(4):1602–1613.

Townsend, P. (1979). *Poverty in the United Kingdom*. Penguin Books, Harmondsworth, Middlesex.

Treibich, R. (2014). Welfare Egalitarianism with Other-regarding Preferences. *Job market paper*.

Warr, P. G. (1982). Pareto Optimal Redistribution and Private Charity. *Journal of Public Economics*, 19:131–138.

Winter, S. (1969). A Simple Remark on the Second Optimality Theorem of Welfare Economics. *Journal of Economic Theory*, 1:99–103.

World Bank (2015). *A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals*. World Bank, Washington, DC.

Zheng, B. (1997). Aggregate Poverty Measures. *Journal of Economic Surveys*, 11(2):123–162.

Zheng, B. (2007). Unit-consistent Poverty Indices. *Economic Theory*, 31(1):113–142.