# ESSAYS ON POLITICAL ECONOMY AND SOCIAL CHOICE

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Economics

2019

By
Minh Tung Le
School of Social Science

# Contents

**Word Count: 15326**

# Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

# Copyright

# Acknowledgement

This thesis cannot be written without the moral and emotional support of my parents, Bich Thi Nguyen and Khai Anh Le, and my sister, Ha Thu Le. Everything that I have achieved in my life is dedicated to them. I would like to express my deepest gratitude to Alejandro Saporiti and Antonio Nicolò for their thoughtful guidance and advise during this difficult journey. In addition, there are my friends, who were of great support in deliberating over my problems, as well as providing happy distraction to rest my mind outside of my research. Finally, to Oanh Oanh Tran, for her love, affection and understanding, which gives a purpose to my every action and a meaning to my one life.

# Chapter 1

# Introduction

Classical theory of social choice is at a large extent concerned with the way preferences or values of individuals are aggregated into the choice of a collective group or society. One of the most effective methods of eliciting one's own preferences is through the act of voting. As a result, the theory of voting itself is the theory of social choice. Along with the evolution of democracy and state institutions in the later half of the twentieth century, the application of social choice theory in political science has become more and more prevalent. Social choice theory provides the sophisticated foundation for the analysis of electoral competition and bargaining processes while political science has become a continual inspiration for many theoretical problems. After all, political economy provides a rich environment for economists to apply their theories to the analysis of human behaviors and political institutions.

The seminal works of Arrow (1950), Black (1958), and Downs (1957) had started a long line of research connecting the field of social choice with the theory of voting and democracy. However, there remains some distance between the theoretical nature of social choice theory and evidence-based practice of political science. Standard game theoretic models suggest that interpersonal comparisons are irrelevant in strategic choice. Although it is a convenient assumption for tractability purposes, political actors are not always an ideal decision-maker, a perfection of rationality or in short, a *homo economicus*. A politician's objective is not only to maximize his own political influence, but also to represent a certain ideal of justice. Voters do not evaluate policy in term of one's own material interest, but on how it affects the society as a whole. By accepting the "theoretical flaws" of political agents, one can make a more precise prediction of political outcome and, more importantly, bring economics closer to a science of human natures and their interaction.

In this book, we contribute to the literature by analyzing and extending some classical models of collective decision-making, relaxing certain standard assumptions in microeconomic theory such as completeness and self-concerned nature. The aim of our research is to show that such irregularities in human behaviors do not only change our prediction of the political outcomes but also open up useful discussions of many well-known results in social choice theory.

In Chapter 2, we extend upon the results of Lindbeck and Weibull (1987) to study distributive politics when voters have not only self-interested preferences, but also other-regarding concerns. We consider a broad family of other-regarding behavior (including *fairness preferences*, *income-dependent altruism*, and *inequality aversion*), for which results on equilibrium existence and optimality have not been yet established. We provide a sufficient condition for smooth and non-smooth payoffs that generalizes Lindbeck and Weibull's condition, and guarantees the existence of a unique Nash equilibrium in pure strategies. In addition, we determine conditions under which the equilibrium results in an income distribution that can be rationalised as the outcome of maximizing a mixture of a "self-regarding utilitarian" social welfare function and society's other-regarding preferences.

In Chapter 3, we relax the differentaiblity assumption and explore the nonemptiness of the core with nonsmooth preferences in classical spatial voting model. Previous analyses on the nonemptyness of the core rely on the assumption that voters' utility functions are pseudo-concave and differentiable to derive the gradient restriction for the core of a voting rule parameterized by an arbitrary quota. To deal with the difficulty caused by the absence of differentiable utility functions, we regard each voter's preference at each kinked point as "unanimous" social preferences of a collective of finite "self", each endowed with a smooth preference over the set of alternatives. By applying the "smooth" analysis to the auxiliary society of selves, we generalize the known gradient restriction for the core of an arbitrary quota rule with any utility profile that satifies the general Lipschitz continuity. We also consider two example of nonsmooth utility representation (inequity aversion and city-block preferences) in the context of simple allocation problem to illustrate the application of our results.

In Chapter 4, we depart from the spatial voting theory and reexamine the social choice problem of Barberà, Sonnenschein and Zhou (1991), where society chooses a subset from a finite set of alternatives (indivisible objects). Each member of society is endowed with an asymmetric and not necessarily complete preference relation over the set of social outcomes (subsets of alternatives). We characterize the family of voting rules satisfying *justifable strategy-proofness*, a notion of incentive compatibility adapted conveniently to accommodate the incompleteness of individual preferences. Under separability, a voting rule is justifiable strategy-proof if and only if it is a voting-by-committees rule. Given the richness of this family, we use Ok's (2002) multi-utility representation for incomplete preferences to further refine our previous characterization, demanding in addition optimality of the social outcome. Our analysis suggests that the structure of the optimal voting-by-committees rules largely depends on how the loss of information due to the presence of preference incompleteness affects the aggregate social welfare.

# Chapter 2

# Distributive Politics with Other-Regarding Preferences

## 2.1 Introduction

Models of political economy, particularly of income redistribution, typically assume that individuals are selfish and care only about their material interests. In the literature on behavioral economics, however, there is mounting evidence that says otherwise, suggesting that people also express concern with the well-being of other individuals in society.[1] The implications of these behavioral studies on individual preferences over payoffs have just started to be examined in political economy. The aim of this paper is to contribute to this new literature by extending the canonical model of distributive politics due to Lindbeck and Weibull (1987) to accommodate a *broad* family of other-regarding behavior, which includes among others inequality aversion (Fehr and Schmidt 1999), fairness concern (Alesina and Angeletos 2005a), and income-dependent altruism (Dimick, Rueda and Stegmueller 2017).

The model laid out in Section 2.2 shares the usual features of probabilistic electoral competition. There are two political parties competing in a single election for the main office. Voters are grouped into different socio-economic groups and have stochastic and policy-independent preferences (ideology) over the parties. The political candidates offer to the electorate a balanced budget

[1]This evidence has been documented in a large number of experimental and neuro-imaging studies, including among many others the work of Fehr and Schmidt (1999), Engelmann and Strobel (2004), Dawes et at. (2007) and  (2012), Tabibnia et al. (2008), Fehr (2009), Almås et al. (2010), Tricomi et al. (2010), Zaki and Mitchell (2011), and Rilling and Sanfey (2011).

redistributive policy from a multidimensional policy space. Voters evaluate these policies taking into account their selfish utility and their ideological bias. In a clear departure from earlier work, in this chapter voters also concern with how these policies affect the well-being of other members of society. To be precise, voters are endowed with an other-regarding utility which is continuous, concave, but not necessarily smooth.[2] Section 2.3 offers a few important examples that match this description. During the campaign, parties choose simultaneously their distributive policies to maximize their expected vote shares, but they care also about voters' other-regarding preferences. The latter implies that the payoff functions of the parties are *not necessarily smooth* on the strategy space.

The main results of the paper are displayed in Section 2.4 and can be summarized as follows. First, the paper generalizes the Lindbeck and Weibull's (1987) sufficient condition for equilibrium existence, adapting it conveniently to accommodate the other-regarding preferences of the electorate and the resulting non-smooth framework described above. This condition, together with the assumptions on the utility functions, namely, continuity and concavity, shape the expected vote share and the parties' payoffs. To start, the gradient of the expected vote share is shown to be monotone decreasing on the differentiable subset of distributive policies (Lemma 2.1).

Since the set of differentiable policy alternatives does not always constitute a convex set, the previous result is not sufficient to prove concavity of the parties' payoff functions. Thus, as a preliminary step it is shown that the expected vote share of each party has a support almost everywhere (Lemma 2.2). Finally, using the fundamental theorem on the support of a concave function, Lemma 2.3 states that the expected vote share is concave on the whole strategy space. This together with the concavity of the other-regarding utilities guarantee that the party payoff functions are concave as well. The existence of Nash equilibrium in pure strategies follows then immediately from the classical Debreu-Glicksberg-Fan's result for games with continuous and quasi-concave payoffs (Theorem 2.1).

Second, the paper studies the properties of the Nash equilibria when the parties hold symmetric electoral goals, meaning that they care equally about winning the election. Using the necessary conditions for the existence of a maximum, Theorem 2.2 characterizes the equilibrium policies of each party, which are shown to

[2]For instance, Fehr and Schmidt's (1999) inequality aversion preferences are not differentiable at the individual's reference point (see equation 2.4).

be unique and the same for both. In addition, the theorem also proves that these policies are "optimal", in the sense that they can be rationalised as the outcome of maximizing a mixture of a "self-regarding utilitarian" social welfare function and society's other-regarding preferences. The optimality result in Lindbeck and Weibull (1987) is derived as a special case under the assumption that society is purely selfish (Corollary 2.1). Finally, third, by strengthening a bit the assumption on the shape of the other-regarding utility, namely, by assuming *strict* concavity, the paper shows in Theorem 2.3 that the uniqueness result stated in Theorem 2.2 holds more generally, and not just under symmetric party motivations, provided that the condition for equilibrium existence is in place.

With regard to the literature most closely related to this article, preferences for redistribution that goes beyond those motivated by the agents' own economic benefits have been studied in Galasso (2003), Alesina and Angeletos (2005a,b), Tyran and Sausgruber (2006), Dhami and al-Nowaihi (2010a,b), Luttens and Valfort (2012), and Flamand (2012). These papers differ from the current work primarily because they focus on the Meltzer and Richard's (1981) median voter framework of redistributive politics, instead of the probabilistic voting (swing voter) model. One of the main limitations of median voter framework is that it cannot capture the multidimensional property of distributive politics. More specifically, in the standard median voter framework, the policy space usually consists of a proportional tax rate and a lump sum transfer as a political equilibrium is unlikely to exist when policy space is defined in more than two dimensions . By using the probabilistic voting model, we allow for policy space in any dimensions which better capture the redistribution problem across different socio-economic groups. Overall, a robust result coming out from this body of research is that the presence of other-regarding preferences leads not only to different predictions concerning the extent of redistribution, but also the link between inequality and redistribution.[3]

In the context of probabilistic electoral competition, to our knowledge the only two articles that incorporates other-regarding preferences into the analysis are Alesina, Cozzi, and Mantovan (2012) and Debowicz, Saporiti, and Wang (2017). The first paper analyzes a dynamic extension of the Lindbeck-Weibull model to explain how different perceptions of fairness of the market outcomes

[3]For example, in the Meltzer-Richard model with social preferences, redistribution depends not only on the mean to median income ratio, but also on the variance of the income distribution.

can lead to different steady states of redistribution and growth. Meanwhile, the second paper, that is, Debowicz et al. (2017), studies the consequences of different distributions of policymaking power over distributive policies and income inequality in the presence of fairness concern. In contrast with the current work, these two papers only focus on fairness, and they do not provide equilibrium existence and optimality results for a broad family of other-regarding preferences, which is the precisely the main objectives of the coming sections.

## 2.2 The Model

There is a society with a continuum of voters divided into $n$ disjoint groups, denoted $N = \{1, 2, \ldots, n\}$, where $n_i \in (0, 1)$ indicates the size of group $i \in N$, and $\sum_{i \in N} n_i = 1$. The initial (finite) gross income of each voter of group $i \in N$ is given by $w_i > 0$. Let $w = \sum_{i \in N} n_i \, w_i$ be the total income of the economy, and denote the set of all possible distributions of $w$ by $Y = \{\mathbf{y} \in \mathbb{R}^N_+ \mid \sum_{i \in N} n_i \, y_i = w\}$.

The preferences of each voter $i \in N$ over $Y$ are additively separable. To be precise, voter $i$'s utility associated with each income distribution $\mathbf{y} \in Y$ is defined as

$$U_i^h(\mathbf{y}) = u_i(y_i) + \alpha_i \, \sigma^h(\mathbf{y}), \tag{2.1}$$

where $u_i : \mathbb{R}_+ \to \mathbb{R}$ is a *self-regarding utility* over disposable income $y_i$, and the function $\sigma^h : \mathbb{R}^N_+ \to \mathbb{R}$ represents voter $i$'s *other-regarding utility*, parameterized by $\alpha_i \in \mathbb{R}_+$, with the index $h$ denoting the specific other-regarding hypothesis under consideration, to be discussed in Section 2.3. These utility functions are assumed to satisfy the following assumptions:

**A1.** $u_i(\cdot)$ is twice continuously differentiable on $\mathbb{R}_+$, with $u_i'(\cdot) > 0$ and $u_i''(\cdot) < 0$.

As the examples in Section 2.3 point out, full differentiability of $\sigma^h(\cdot)$ is not always guaranteed under the different models of other-regarding preferences that this paper aims to accommodate. A case in point is inequality aversion, where the utility has a kink and it is not differentiable at the individual's reference point (own payoff). To deal with these cases, assume that the other-regarding utility verifies the following assumption:

**A2.** $\sigma^h(\cdot)$ is continuous and concave on $Y$.

**A3.** $\sigma^h(\cdot)$ is twice continuously differentiable almost everywhere except possibly on a subset of Lebesgue measure zero.

Assumption **A2** and **A3** provide certain implications over the functional form of $\sigma^h(.)$. First of all, the concavity of $\sigma^h(.)$ implies that generally, voters enjoy an equitable distribution over the extreme alternative. Secondly, as we only require Lipschitz continuity of the utility function, it allows us to extend our analysis to a wide range of reference dependent preferences with inequity aversion being a special case.

Moving to the political setup, there are two political parties, indexed by $C = A, B$, that compete in a single election proposing simultaneously a tax-and-transfer distributive policy $\mathbf{x}_C \in X = \{\mathbf{x}'_C \in \mathbb{R}^N \mid \sum_{i\in N} n_i\, x'_{iC} = 0$ and $\forall i \in N,\ x'_{iC} \geq -w_i\}$. Given that the initial income of each group is held fixed during the analysis, define accordingly from **A3** a subset of policies $X_d \subset X$ where the party payoff functions (yet to be defined) are smooth, with $\overline{X}_d = X \backslash X_d$ denoting the subset where they are not. This notation will be used in Section 2.4 along the proofs of the main results.

A voter in group $i \in N$ votes for party $A$ if $U_i^h(\mathbf{x}_A) \geq U_i^h(\mathbf{x}_B) + \theta_i$,[4] where $\theta_i \in \mathbb{R}$ denotes voter $i$'s policy-independent preference bias towards party $B$, drawn from a twice continuously differentiable distribution function $F_i$, with density $f_i$ positive everywhere over the interval that includes all possible values of the utility differences $t_i^h(\mathbf{x}_A,\, \mathbf{x}_B) = U_i^h(\mathbf{x}_A) - U_i^h(\mathbf{x}_B)$.[5] The (expected) vote share of party $A$ is given by $v_A^h(\mathbf{x}_A, \mathbf{x}_B) = \sum_{i\in N} n_i\, F_i(U_i^h(\mathbf{x}_A) - U_i^h(\mathbf{x}_B))$. Assuming no voter abstention, party $B$'s vote share is $v_B^h = 1 - v_A^h$.

The payoff functions of the parties, viz. $\Pi_C^h$, express the interests of the politicians, who campaign to maximize their vote share (expected plurality). In addition, the payoffs reflect the views of regular party members, who see the party as a vehicle to promote not just their own interest, but also the well-being of others in society. Formally, the payoff function of party $C$ is defined as $\Pi_C^h(\mathbf{x}_A, \mathbf{x}_B) = v_C^h(\mathbf{x}_A, \mathbf{x}_B) + \alpha_C\, \sigma^h(\mathbf{x}_C)$, where $\alpha_C \in [0, \infty)$ is the *relative* value that party $C$ assigns to other-regarding concerns.[6]

[4]To save on notation and given that the initial income $w_i$ is fixed, the utility $U_i^h$ is written simply as a function of $\mathbf{x}_C$, instead of the disposable incomes $\mathbf{y}_C = (y_{iC})_{i\in N} \in Y$, where $y_{iC} = w_i + x_{iC}$. When there is no risk of confusion, the same notation is adopted for other functions that also depend on $\mathbf{y}_C$.

[5]Instead of being additive, the preference bias can be a multiplicative factor on the utility of policy, implying that party $A$ is preferred by $i$ if $U_i^h(\mathbf{x}_A) \geq \theta_i\, U_i^h(\mathbf{x}_B)$. Given that the logarithm of $U_i^h(\cdot)$ is also a utility, the results obtained for the additive case extend directly to the multiplicative model.

[6]Alternatively, $\alpha_C$ can be seen in some cases as the reputation cost for the party of campaigning on distributive policies perceived by the electorate as "socially insensible" (i.e., the

Let $\mathcal{G}^h = (X, \Pi^h_C)_{C=A,B}$ denote the ***distributive election game*** determined by the model sketched above. The timing of $\mathcal{G}^h$ is as follows. First, parties $A$ and $B$ choose simultaneously and non-cooperatively $\mathbf{x}_A$ and $\mathbf{x}_B$, respectively. At this stage, parties know the initial income of the groups, voters' preferences over the income distributions, and the group-specific cumulative distributions of the preference bias. Second, the actual values of $\theta_i$ are realized. Third, voters cast their vote for one of the parties. Fourth, plurality rule determines the winning party (with ties broken by a fair lottery) and its policy platform is implemented. Finally, fifth, parties and voters receive their payoffs.

A pure-strategy Nash equilibrium of $\mathcal{G}^h = (X, \Pi^h_C)_{C=A,B}$ is a policy profile $(\hat{\mathbf{x}}_A, \hat{\mathbf{x}}_B) \in X \times X$ such that $\Pi^h_C(\hat{\mathbf{x}}_C, \hat{\mathbf{x}}_{-C}) \geq \Pi^h_C(\mathbf{x}', \hat{\mathbf{x}}_{-C})$ for all $\mathbf{x}' \in X$ and $C = A, B$, where the index $-C$ denotes $B$ if $C = A$ and $A$ if $C = B$.

## 2.3 Other-Regarding Preferences

This section offers a few important examples of other-regarding behavior that fit well into the model of Section 2.2. Consider first Alesina and Angeletos' (2005a) *fairness preferences* (FP) hypothesis. The distinctive feature of this hypothesis is that individuals distinguish between *fair* and *unfair* income inequality, and they express dislike and concern only for the second. To be more precise, suppose the initial income of voter $i \in N$ is given by $w_i = e_i + \eta_i$, where $e_i$ denotes his fair (earned) income, received in compensation for talent and effort, and $\eta_i$ indicates his unfair (unearned) income, obtained through lucky or illicit transactions. Assume $\eta_i$ is distributed independently from $e_i$ with zero mean.

In the presence of fairness concerns, the other-regarding utility corresponding to any income distribution $\mathbf{y} \in Y$ takes the form

$$\sigma^{FP}(\mathbf{y}) = -\sum_{i \in N} n_i \, (y_i - e_i)^2, \tag{2.2}$$

which captures that only unfair income comes at a utility cost to the individuals.

A second hypothesis of other-regarding behavior corresponds to the model proposed by Dimick, Rueda and Stegmueller (2017), named *income-dependent*

cost of building the image of being a "nasty party" that only cares about the privileged few and not the many, as the British Conservative Prime Minister, Theresa May, put it in her 2002 party conference speech).

*altruism* (IDA). The main assumption is that individuals are concerned with aggregate social welfare. To be concrete, under this hypothesis the other-regarding utility of any income distribution $\mathbf{y} \in Y$ takes the form of the standard utilitarian social welfare function,

$$\sigma^{IDA}(\mathbf{y}) = \sum_{i \in N} n_i \, u_i(y_i), \tag{2.3}$$

which is the sum of individuals' self-regarding utilities, each weighted by the group size.

Finally, the third hypothesis discussed here is the (reference-dependent) *inequality aversion* (IA) model of Fehr and Schmidt (1999). The key feature of it is that individuals evaluate inequality differently depending on the position of their own payoff relative to the others. For any $\mathbf{y} \in Y$, the other-regarding utility of voter $i$ is

$$\sigma^{IA}(\mathbf{y}) = -\gamma \sum_{j \neq i} n_j \, \max\{y_j - y_i, 0\} - \beta \sum_{j \neq i} n_j \, \max\{y_i - y_j, 0\}, \tag{2.4}$$

where $\beta \leq \gamma$ and $\beta \in [0, 1)$.[7] The first (resp., second) term in the right-hand side of equation (2.4) represents group $i$'s disadvantageous (resp., advantageous) inequality, weighted by $\gamma$ (resp., $\beta$). The assumption is that individuals are more selfish than altruistic, and consequently that they are more concerned with disadvantageous inequality.[8]

Notice that the examples of other-regarding preferences given above are associated with continuous and concave utility functions, which are differentiable everywhere except possibly on a set of points of Lebesgue measure zero. There are other examples of other-regarding behavior relevant for distributive politics which also share the properties of assumptions **A2** and **A3**, including maximin and quasi-maximin preferences, efficiency concerns, Bolton and Ockenfels' (2000) inequality aversion model, etc.[9] The next section explores equilibrium existence

[7] Dhami and al-Nowaihi (2010a) and (2010b) consider a generalization of Fehr and Schmidt's (1999) model where payoff comparisons are not made in terms of monetary payoffs, but in utility terms. A drawback of that model is that the other-regarding utility is not necessarily concave.

[8] Inequality aversion preferences are self-centered, because individuals use their payoff as a reference point with which everyone else is compared to. However, people are not concerned with inequality per se. This stands in opposition with experimental evidence, which shows that in simple distribution games people also consider differences among others in their utility functions (Engelmann and Strobel 2004).

[9] See Engelmann and Strobel (2007), Alesina and Giuliano (2010), Clark and D'Ambrosio (2015), Dhami (2016) and the references for alternative theories of redistribution preferences.

and optimality within the class of social preferences that satisfy these restrictions and guarantee the concavity of the parties' conditional payoff functions.

## 2.4 Results

To start the analysis, define below a condition, denoted $\mathbb{C}^h$, that generalizes Lindbeck and Weibull's (1987) sufficient condition (see the discussion at the end of the paper), conveniently adapted for the framework laid out in Section 2.2. Fix any $\mathbf{x}_{-C} \in X$, we can write $t_i^h(\mathbf{x}_C)$ instead of $t_i^h(\mathbf{x}_C,\, \mathbf{x}_{-C})$.

**Condition $\mathbb{C}^h$:** For all $i \in N$,

$$\inf_{\mathbf{x},\, \hat{\mathbf{x}} \in X_d} \left( \frac{f_i(t_i^h(\mathbf{x})) - f_i(t_i^h(\hat{\mathbf{x}}))}{f_i(t_i^h(\mathbf{x}))} \cdot \frac{\sum_{j\in N} \frac{\partial U_i^h(\hat{\mathbf{x}})}{\partial x_j} \cdot (x_j - \hat{x}_j)}{\sum_{j\in N} \left( \frac{\partial U_i^h(\mathbf{x})}{\partial x_j} - \frac{\partial U_i^h(\hat{\mathbf{x}})}{\partial x_j} \right) \cdot (x_j - \hat{x}_j)} \right) \geq -1.$$

The next three propositions illustrate how $\mathbb{C}^h$ together with **A1**, **A2** and **A3** shape the conditional payoffs of the parties. Beginning with the gradient of the expected vote shares, these are shown to be monotone decreasing on the differentiable subset of distributive policies, a result that follows immediately from condition $\mathbb{C}^h$. A function $f : X \to \mathbb{R}^N$ is said to be monotone decreasing on $X$ if for all $x^1, x^2 \in X$, we have $[f(x^1) - f(x^2)](x^1 - x^2) \leq 0$.

**Lemma 2.1** *Suppose assumptions* **A1**–**A2** *hold. Under condition $\mathbb{C}^h$, for each $C = A, B$ and all $\mathbf{x}_{-C} \in X$, the gradient $\nabla v_C^h(\,\cdot,\, \mathbf{x}_{-C})$ is monotone decreasing on $X_d$.*

**Proof.** Fix any $\mathbf{x}_{-C} \in X$. The gradient of party $C$'s expected vote share $\nabla v_C^h(\,\cdot,\, \mathbf{x}_{-C})$ is monotone decreasing on $X_d$ if for all $\mathbf{x}^1, \mathbf{x}^2 \in X_d$,

$$\left[\nabla v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) - \nabla v_C^h(\mathbf{x}^2, \mathbf{x}_{-C})\right] \cdot (\mathbf{x}^1 - \mathbf{x}^2) \leq 0. \tag{2.5}$$

Recall that the conditional vote share $v_C^h(\mathbf{x}_C, \mathbf{x}_{-C}) = \sum_{i\in N} n_i\, v_{iC}^h(\mathbf{x}_C, \mathbf{x}_{-C})$, where $v_{iC}^h(\mathbf{x}_C, \mathbf{x}_{-C}) = F_i(U_i^h(\mathbf{x}_C) - U_i^h(\mathbf{x}_{-C}))$. Thus, inequality (2.5) holds if for all $i \in N$,

$$\sum_{j\in N}\left[f_i(t_i^h(\mathbf{x}^1))\frac{\partial U_i(\mathbf{x}^1)}{\partial x_j}-f_i(t_i^h(\mathbf{x}^2))\frac{\partial U_i(\mathbf{x}^2)}{\partial x_j}\right](x_j^1-x_j^2)\leq 0$$

$$\iff \frac{f_i(t_i^h(\mathbf{x}^1))-f_i(t_i^h(\mathbf{x}^2))}{f_i(t_i^h(\mathbf{x}^1))}\cdot\frac{\sum_{j\in N}\frac{\partial U_i^h(\mathbf{x}^2)}{\partial x_j}(x_j^1-x_j^2)}{\sum_{j\in N}\left[\frac{\partial U_i^h(\mathbf{x}^1)}{\partial x_j}-\frac{\partial U_i^h(\mathbf{x}^2)}{\partial x_j}\right](x_j^1-x_j^2)}\geq -1,$$

which is implied by condition $\mathbb{C}^h$. ∎

Due to the presence of non-differentiable points, the policy subset $X_d$ is not necessarily convex. Hence, Lemma 2.1 is not enough to prove the concavity of $v_C^h(\,\cdot,\,\mathbf{x}_{-C})$. To do so, we need another preliminary result, which ensures that the expected vote share of each party has a support on $X_d$. A function $f : X \to \mathbb{R}$ has a support at an alternative $x^1 \in X$ if there exists a vector $\mathbf{a} \in \mathbb{R}^N$ such that for any $x^2 \in X$, we have $f(x^2) \leq f(x^1) + \mathbf{a}\cdot(x^2 - x^1)$.

**Lemma 2.2** *Suppose assumptions* **A1**–**A3** *hold. Under condition* $\mathbb{C}^h$*, for each* $C = A, B$ *and all* $\mathbf{x}_{-C} \in X$*, the expected vote share* $v_C^h(\,\cdot,\,\mathbf{x}_{-C})$ *has a support at each* $\mathbf{x} \in X_d$*.*

**Proof.** Fix any $\mathbf{x}_{-C} \in X$. The expected vote share $v_C^h(\,\cdot,\,\mathbf{x}_{-C})$ has a support at $\mathbf{x}^1 \in X_d$ if there exists a vector $\mathbf{a}(\mathbf{x}^1) \in \mathbb{R}^N$ such that for any $\mathbf{x}^2 \in X_d$,

$$v_C^h(\mathbf{x}^2,\mathbf{x}_{-C}) \leq v_C^h(\mathbf{x}^1,\mathbf{x}_{-C}) + \mathbf{a}(\mathbf{x}^1)\cdot(\mathbf{x}^2-\mathbf{x}^1). \tag{2.6}$$

Consider any $\mathbf{x}'$, $\mathbf{x}'' \in X$, and let $S(\mathbf{x}',\mathbf{x}'') = \{\delta\mathbf{x}' + (1-\delta)\mathbf{x}'' \in X$, with $\delta \in (0,1)\}$. Fix *any* $\mathbf{x}^2 \in X_d$. There are three cases to study.

**Case 1.** Suppose $S(\mathbf{x}^1,\mathbf{x}^2) \cap \overline{X}_d = \emptyset$. Then, the function $v_C^h(\,\cdot\,,\mathbf{x}_{-C})$ is differentiable on $S(\mathbf{x}^1,\mathbf{x}^2)$. Taking the gradient $\nabla v_C^h(\mathbf{x}^1,\mathbf{x}_{-C})$ as the support vector $\mathbf{a}(\mathbf{x}^1)$, inequality (2.6) holds because $S(\mathbf{x}^1,\mathbf{x}^2)$ is open and convex, and consequently Lemma 2.1 implies that $v_C^h(\,\cdot\,,\mathbf{x}_{-C})$ is concave on it.

**Case 2.** Assume $S(\mathbf{x}^1,\mathbf{x}^2)\cap\overline{X}_d = \{\mathbf{z}^k,\ k = 1,\ldots,K\}$, where $K$ is a finite positive integer. For each $k$, let $\mathbf{z}^k = \lambda^k\,\mathbf{x}^1 + (1-\lambda^k)\,\mathbf{x}^2$ for some $\lambda^k \in (0,1)$. Without loss of generality, assume $\lambda^1 < \ldots < \lambda^K$. Consider the open and convex subsets $S(\mathbf{x}^1,\mathbf{z}^1)$, $S(\mathbf{z}^1,\mathbf{z}^2),\ldots,S(\mathbf{z}^K,\mathbf{x}^2)$. The expected vote share $v_C^h(\,\cdot\,,\mathbf{x}_{-C})$ is smooth on each of these subsets. Using the argument of Case 1 on the first subset $S(\mathbf{x}^1, \mathbf{z}^1)$, it follows from (2.6) that

$$v_C^h(\mathbf{z}^1, \mathbf{x}_{-C}) \leq v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) + \nabla v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) \cdot (\mathbf{z}^1 - \mathbf{x}^1). \tag{2.7}$$

Applying the same reasoning on the second subset, i.e., $S(\mathbf{z}^1, \mathbf{z}^2)$, and invoking the continuity of the expected vote share $v_C^h(\,\cdot\,, \mathbf{x}_{-C})$, we have that

$$v_C^h(\mathbf{z}^2, \mathbf{x}_{-C}) \leq v_C^h(\mathbf{z}^1, \mathbf{x}_{-C}) + \lim_{\delta \to 0} \nabla v_C^h(\delta \mathbf{z}^2 + (1-\delta)\mathbf{z}^1, \mathbf{x}_{-C}) \cdot (\mathbf{z}^2 - \mathbf{z}^1). \tag{2.8}$$

Notice in the above inequality that since $\mathbf{z}^1 \in \overline{X}_d$, $\lim_{\delta \to 0} \nabla v_C^h(\delta \mathbf{z}^2 + (1-\delta)\mathbf{z}^1, \mathbf{x}_{-C})$ represents the superdifferential $\partial_S v_C^h(\mathbf{z}^1, \mathbf{x}_{-C})$ of $v_C^h(\,\cdot\,, \mathbf{x}_{-C})$ at $\mathbf{z}^1$, and that (2.8) holds for each supergradient vector in $\partial_S v_C^h(\mathbf{z}^1, \mathbf{x}_{-C})$.[10] Adding up (2.7) and (2.8) and using the fact that by Lemma 2.1,

$$\Big[\nabla v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) - \lim_{\delta \to 0} \nabla v_C^h(\delta \mathbf{z}^2 + (1-\delta)\mathbf{z}^1, \mathbf{x}_{-C})\Big] \cdot (\mathbf{z}^2 - \mathbf{z}^1) \geq 0,$$

it follows that

$$v_C^h(\mathbf{z}^2, \mathbf{x}_{-C}) \leq v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) + \nabla v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) \cdot (\mathbf{z}^2 - \mathbf{x}^1). \tag{2.9}$$

Finally, the desired result is obtained by repeating the previous argument over all the remaining subsets, which proves that inequality (2.6) holds strictly on $S(\mathbf{x}^1, \mathbf{x}^2)$ with support vector $\mathbf{a}(\mathbf{x}^1) = \nabla v_C^h(\mathbf{x}^1, \mathbf{x}_{-C})$.

**Case 3.** Suppose $S(\mathbf{x}^1, \mathbf{x}^2) \cap \overline{X}_d = \cup_{k=1}^K A^k$, where each $A^k$ is a closed and convex subset, and $K$ is a finite positive integer. If $A^k$ is a singleton for every $k$, then this case coincides with Case 2. Otherwise, there must exist some $k$ and $\underline{\delta}$, $\overline{\delta} \in (0,1)$ such that the subset $A^k = \{\delta \mathbf{x}^1 + (1-\delta)\mathbf{x}^2 \in \overline{X}_d$, with $\delta \in [\underline{\delta}, \overline{\delta}]\}$. By **A3**, for all $\epsilon > 0$ there exists $\hat{\mathbf{x}}^2 \in B_\epsilon(\mathbf{x}^2)$ such that $S(\mathbf{x}^1, \hat{\mathbf{x}}^2) \cap \overline{X}_d$ is a finite set. Using the argument of Case 2, note that equation (2.6) holds strictly on $S(\mathbf{x}^1, \hat{\mathbf{x}}^2)$, with support vector $\mathbf{a}(\mathbf{x}^1) = \nabla v_C^h(\mathbf{x}^1, \mathbf{x}_{-C})$. Applying the continuity of the function $v_C^h(\,\cdot\,, \mathbf{x}_{-C})$ gives

$$v_C^h(\mathbf{x}^2, \mathbf{x}_{-C}) \leq v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) + \nabla v_C^h(\mathbf{x}^1, \mathbf{x}_{-C}) \cdot (\mathbf{x}^2 - \mathbf{x}^1), \tag{2.10}$$

[10] Recall that the superdifferential of a function $f : X \subset \mathbb{R}^N \to \mathbb{R}$ at $\mathbf{x} \in X$ is the set of vectors $\partial_S f(\mathbf{x}) = \big\{\mathbf{a} \in \mathbb{R}^N \mid f(\hat{\mathbf{x}}) \leq f(\mathbf{x}) + \mathbf{a} \cdot (\hat{\mathbf{x}} - \mathbf{x})$, for all $\hat{\mathbf{x}} \in X\big\}$.

which concludes the argument for $A^k$ and completes the proof that $v_C^h(\,\cdot,\, \mathbf{x}_{-C})$ has a support at $\mathbf{x}$. ■

We are finally ready to show the concavity of the expected vote share.

**Lemma 2.3** *Suppose assumptions* **A1**–**A3** *hold. Under condition* $\mathbb{C}^h$*, for each* $C = A, B$ *and all* $\mathbf{x}_{-C} \in X$*, the expected vote share* $v_C^h(\cdot, \mathbf{x}_{-C})$ *is concave on* $X$*.*

**Proof.** Fix any $\mathbf{x}_{-C} \in X$. By the fundamental theorem on the support of a concave function, $v_C^h(\cdot, \mathbf{x}_{-C})$ is concave on $X$ if and only if it has support at each interior point of $X$. By Lemma 2.2, $v_C^h(\cdot\,, \mathbf{x}_{-C})$ has support on $X_d$. The rest of the proof is based on the following two claims.

**Claim 1** *For each* $\mathbf{x} \in X_d$*, the support vector of* $v_C^h(\cdot, \mathbf{x}_{-C})$ *at* $\mathbf{x}$ *holds for all* $\mathbf{x}^0 \in \overline{X}_d$*.*

Fix $\mathbf{x} \in X_d$ and consider any $\mathbf{x}^0 \in \overline{X}_d$. By **A3**, for all $\epsilon > 0$ there exists $\mathbf{x}' \in B_\epsilon(\mathbf{x}^0) \cap X_d$. By Lemma 2.2, the function $v_C^h(\cdot\,, \mathbf{x}_{-C})$ has support over $X_d$, meaning that $v_C^h(\mathbf{x}', \mathbf{x}_{-C}) \le v_C^h(\mathbf{x}, \mathbf{x}_{-C}) + \nabla v_C^h(\mathbf{x}, \mathbf{x}_{-C}) \cdot (\mathbf{x}' - \mathbf{x})$. Since $v_C^h(\cdot\,, \mathbf{x}_{-C})$ is continuous on $X$, taking the limit of the previous inequality as $\mathbf{x}' \to \mathbf{x}^0$ gives

$$v_C^h(\mathbf{x}^0, \mathbf{x}_{-C}) \le v_C^h(\mathbf{x}, \mathbf{x}_{-C}) + \nabla v_C^h(\mathbf{x}, \mathbf{x}_{-C}) \cdot (\mathbf{x}^0 - \mathbf{x}),$$

which provides the desired result.

**Claim 2** $v_C^h(\cdot\,, \mathbf{x}_{-C})$ *has support at each* $\mathbf{x}^0 \in \overline{X}_d$*.*

Fix $\mathbf{x}^0 \in \overline{X}_d$. By **A3**, for sufficient small $\epsilon > 0$, there exists $\mathbf{x}' \in B_\epsilon(\mathbf{x}^0) \cap X_d$ such that $\{\delta \mathbf{x}' + (1-\delta)\mathbf{x}^0 : \delta \in (0,1)\} \subset X_d$. Using the support of $v_C^h(\cdot\,, \mathbf{x}_{-C})$ at $\mathbf{x}'$ and taking the limit as $\mathbf{x}' \to \mathbf{x}^0$, for all $\mathbf{x} \in X$,

$$v_C^h(\mathbf{x}, \mathbf{x}_{-C}) \le v_C^h(\mathbf{x}^0, \mathbf{x}_{-C}) + \lim_{\delta \to 0} \nabla v_C^h(\delta \mathbf{x}' + (1-\delta)\mathbf{x}^0, \mathbf{x}_{-C}) \cdot (\mathbf{x} - \mathbf{x}^0).$$

Hence, $v_C^h(\cdot\,, \mathbf{x}_{-C})$ has support at $\mathbf{x}^0$. ■

The next theorem generalizes Lindbeck and Weibull's (1987) existence result for probabilistic electoral competition, establishing the existence of a Nash equilibrium in pure strategies when voters' other-regarding concerns permit a large degree of preference interdependence, which may imply that the payoff functions of the parties are not smooth on the strategy space. The proof follows immediately from Debreu-Glicksberg-Fan's existence result. Indeed, note first that the strategy space $X$ is non-empty, compact, and convex.[11] Second, each payoff function $\Pi^h_C(\mathbf{x}_A, \mathbf{x}_B)$ is continuous on $(\mathbf{x}_A, \mathbf{x}_B) \in X \times X$. Finally, third, Lemma 2.3 together with assumption **A2** guarantee that the conditional payoff functions $\Pi^h_C(\,\cdot\,, \mathbf{x}_{-C})$ are concave in the party's own strategy $\mathbf{x}_C$.

**Theorem 2.1 (Existence)** *Suppose assumptions* **A1**–**A3** *hold. Under condition* $\mathbb{C}^h$*, the election game* $\mathcal{G}^h = (X, \Pi^h_C)_{C=A,B}$ *has a pure strategy Nash equilibrium.*

The result stated above guarantees the existence of Nash equilibrium in pure strategies in a broad family of income redistribution games. This includes games with and without social preferences, with voters and parties displaying several patterns of other-regarding behavior, and also with *symmetric* (i.e., $\alpha_A = \alpha_B$) and *asymmetric* (i.e., $\alpha_A \neq \alpha_B$) other-regarding concerns in the parties' payoff functions.

While the theorem constitutes an essential part of the equilibrium analysis, existence *per se* is only the first step. To use the model for predictive purposes requires being able to spell the properties of the policies played in equilibrium. The rest of this section deals with this matter. In particular, it focuses on the conditions under which the equilibrium is unique and it results in an "optimal" after-tax income distribution, in the sense that it can be rationalised as the outcome of maximizing a "sound" social welfare function.

Define the weighted (self-regarding) utilitarian social welfare function as $W(\mathbf{x}) = \sum_{i\in N} n_i\, f_i(0)\, u_i(w_i{+}x_i)$. Let $X^0 = \{\mathbf{x} \in X : w_i{+}x_i > 0 \text{ for all } i \in N\}$.[12] The next result yields the following equilibrium characterization.

**Theorem 2.2 (Characterization)** *Suppose assumptions* **A1**–**A3** *hold. Let* $(\hat{\mathbf{x}}_A, \hat{\mathbf{x}}_B) \in X^0 \times X^0$ *be the Nash equilibrium of* $\mathcal{G}^h$*. If* $\alpha_A = \alpha_B \equiv \bar{\alpha}$*, then* $\hat{\mathbf{x}}_A = \hat{\mathbf{x}}_B \equiv \hat{\mathbf{x}}$*, and*

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\in X^0}\; W(\mathbf{x}) + \Sigma^h(\mathbf{x}), \tag{2.11}$$

[11] In this paper, the policy space $X$ is determined by the resource constraint and the groups' nonnegative income constraints. However, the proof of Theorem 2.1 applies more generally, provided that non-emptiness, compactness, and convexity are preserved. That includes other typical restrictions on $X$, such as non-income sorting among different socio-economic groups (cf. Debowicz et al. 2017).

[12] Lindbeck and Weibull (1987) also assume that each voter's disposable income is strictly positive.

*where* $\Sigma^h(\hat{\mathbf{x}}) = \sum_{i\in N} \psi_i\, \sigma^h(\hat{\mathbf{x}}),$ *with* $\psi_i \equiv n_i\, f_i(0)\, \alpha_i + \bar{\alpha}.$

The proof is similar to that of Lindbeck and Weibull (1987), taking into account the nonsmoothness of voters' utility functions.

**Proof.** Fix the equilibrium strategy $\hat{\mathbf{x}}_{-C}$ and consider the constrained optimization problem of party $C$, which consists in maximizing with respect to $\mathbf{x}_C \in \mathbb{R}^N$ the function $\Pi^h_C(\mathbf{x}_C, \hat{\mathbf{x}}_{-C}) = v^h_C(\mathbf{x}_C, \hat{\mathbf{x}}_{-C}) + \alpha_C\, \sigma^h(\mathbf{x}_C)$, subject to $\sum_{i\in N} n_i x_{iC} = 0$ and $w_i + x_{iC} > 0$, all $i \in N$. By assumption **A3**, $\Pi^h_C(\,\cdot\,, \hat{\mathbf{x}}_{-C})$ is is twice continuously differentiable almost everywhere on $X$. Using the Karush-Kuhn-Tucker optimality conditions, it follows that a necessary condition for a maximum requires that for each group $i \in N$, there exists supergradient vector $p(\hat{\mathbf{x}}_C) \in \partial_S \sigma^h(\hat{\mathbf{x}}_C)$ such that

$$n_i f_i(t^h_i(\hat{\mathbf{x}}_C, \hat{\mathbf{x}}_{-C})) \left( \frac{\partial u_i(w_i + \hat{x}_{iC})}{\partial x_{iC}} + \alpha_i p(\hat{\mathbf{x}}_C) \cdot \mathbf{i} \right) + \alpha_C p(\hat{\mathbf{x}}_C) \cdot \mathbf{i} + n_i \lambda_C = 0, \quad (2.12)$$

where $\lambda_C \geq 0$ is the Lagrange multiplier on the party's resource constraint and $\mathbf{i} \in \mathbb{R}^N$ is the unit vector in the direction of group $i$'s income.

Using the above expression for both parties and after some algebraic manipulation the following condition characterizes the equilibrium policies:

$$\frac{\lambda_A}{\lambda_B} = \frac{n_i f_i(t^h_i(\hat{\mathbf{x}}_A, \hat{\mathbf{x}}_B)) \left( \frac{\partial u_i(w_i + \hat{x}_{iA})}{\partial x_{iA}} + \alpha_i p(\hat{\mathbf{x}}_A) \cdot \mathbf{i} \right) + \alpha_A p(\hat{\mathbf{x}}_A) \cdot \mathbf{i}}{n_i f_i(t^h_i(\hat{\mathbf{x}}_B, \hat{\mathbf{x}}_A)) \left( \frac{\partial u_i(w_i + \hat{x}_{iB})}{\partial x_{iB}} + \alpha_i p(\hat{\mathbf{x}}_B) \cdot \mathbf{i} \right) + \alpha_B p(\hat{\mathbf{x}}_B) \cdot \mathbf{i}}. \quad (2.13)$$

It is easy to see that if $\alpha_A = \alpha_B$, then a solution to (2.13) is given by $\hat{\mathbf{x}}_A = \hat{\mathbf{x}}_B$ and $\lambda_A = \lambda_B$. In fact, there is no other solution with $\hat{\mathbf{x}}_A = \hat{\mathbf{x}}_B$ and $\lambda_A \neq \lambda_B$. Therefore, any other critical point must be such that $\hat{\mathbf{x}}_A \neq \hat{\mathbf{x}}_B$. Without loss of generality, let $\hat{x}_{iA} < \hat{x}_{iB}$ for some $i \in N$. By the resource constraint, there exists a $j \in N$ such that $\hat{x}_{jA} > \hat{x}_{jB}$. By the strict concavity of the self-regarding utility, $\frac{\partial u_i(w_i+\hat{x}_{iA})}{\partial x_{iA}} > \frac{\partial u_i(w_i+\hat{x}_{iB})}{\partial x_{iB}}$. By the definition of supergradient vector for non-smooth concave functions, $p(\hat{\mathbf{x}}_A) \cdot \mathbf{i} \geq \frac{\sigma^h(\hat{\mathbf{x}}_B) - \sigma^h(\hat{\mathbf{x}}_A)}{\hat{x}_{iB} - \hat{x}_{iA}} \geq p(\hat{\mathbf{x}}_B) \cdot \mathbf{i}$. Since $f_i(t^h_i(\hat{\mathbf{x}}_A, \hat{\mathbf{x}}_B))$ are positive and the same for both parties, equation (2.13) implies that $\lambda_A > \lambda_B$. Repeating the argument for group $j \in N$, with $\frac{\partial u_j(w_j+\hat{x}_{jA})}{\partial x_{jA}} < \frac{\partial u_j(w_j+\hat{x}_{jB})}{\partial x_{jB}}$ and $p(\hat{\mathbf{x}}_A) \cdot \mathbf{j} \leq p(\hat{\mathbf{x}}_B) \cdot \mathbf{j}$, it follows that $\lambda_A < \lambda_B$, a contradiction. Hence, $\hat{\mathbf{x}}_A = \hat{\mathbf{x}}_B$ is the only solution to (2.13), and consequently $f_i(t^h_i(\hat{\mathbf{x}}_A, \hat{\mathbf{x}}_A)) = f_i(0)$.

Finally, equation (2.11) follows by applying the Karush-Kuhn-Tucker optimality conditions to the object function $W(\mathbf{x}) + \Sigma^h(\mathbf{x})$, and realizing that the resulting necessary conditions for a maximum coincide with the expression in (2.12). ∎

The previous theorem shows that when parties are symmetric, in the sense that they value power similarly, their equilibrium distributive policies coincide. This result is driven by the fact that parties' constraint optimization problems share the same necessary conditions. The theorem points out that these conditions also characterize the solution of the social planner's wealth allocation problem, provided that its objective consists in maximizing some weighted (self-regarding) utilitarian social welfare function plus an "aggregate" of individuals' and parties' other-regarding preferences.

In the special case of a purely selfish society, Theorem 2.2 offers as a corollary the following well-known result due to Lindbeck and Weibull (1987).

**Corollary 2.1 (Lindbeck-Weibull)** *Under the hypotheses of Theorem 2.2, if* $\alpha_i = \bar{\alpha} = 0$*, then* $\hat{\mathbf{x}}$ *maximizes the weighted (self-regarding) utilitarian social welfare function, i.e.,*

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\in X^0} W(\mathbf{x}).$$

Notice that the result in Theorem 2.2 offers a unique equilibrium prediction for the distributive election game. This is actually preserved in a more general family of distribution games which are not necessarily symmetric. To elaborate, let us assume that the other-regarding utility satisfies the following stronger version of assumption **A2**, as is the case with Alesina and Angeletos' (2005a) fairness preferences and Dimick, Rueda and Stegmueller's (2017) income-dependent altruism.

**A2***. $\sigma^h(\cdot)$ is continuous and *strictly* concave on $Y$.

The assumption above in conjunction with the other conditions already employed allow to state the last result of the paper.

**Theorem 2.3 (Uniqueness)** *If condition* $\mathbb{C}^h$ *and assumptions* **A1**, **A2***, *and* **A3** *hold, then the equilibrium of* $\mathcal{G}^h = (X, \Pi^h_C)_{C=A,B}$ *is unique.*

**Proof.** Suppose, by contradiction, that $(\mathbf{x}'_A, \mathbf{x}'_B)$ and $(\mathbf{x}''_A, \mathbf{x}''_B)$ are two Nash equilibria of $\mathcal{G}^h = (X, \Pi^h_C)_{C=A,B}$. Without loss of generality, let $\mathbf{x}'_A \neq \mathbf{x}''_A$. If $(\mathbf{x}''_A, \mathbf{x}'_B)$ is a Nash equilibrium, then $\Pi^h_A(\mathbf{x}'_A, \mathbf{x}'_B) = \Pi^h_A(\mathbf{x}''_A, \mathbf{x}'_B)$. By assumption **A2***

and Lemma 2.3, for each $C = A, B$ and all $\mathbf{x}_{-C} \in X$, $\Pi^h_C(\,\cdot\,, \mathbf{x}_{-C})$ is strictly concave on $X$. Thus, for all $\delta \in (0,1)$, $\Pi^h_A(\mathbf{x}^\delta_A, \mathbf{x}'_B) > \Pi^h_A(\mathbf{x}'_A, \mathbf{x}'_B)$, with $\mathbf{x}^\delta_A = \delta\mathbf{x}'_A + (1-\delta)\mathbf{x}''_A$, contradicting that $(\mathbf{x}'_A, \mathbf{x}'_B)$ is a Nash equilibrium. Therefore, $\Pi^h_A(\mathbf{x}'_A, \mathbf{x}'_B) > \Pi^h_A(\mathbf{x}''_A, \mathbf{x}'_B)$; and by the same token, $\Pi^h_A(\mathbf{x}''_A, \mathbf{x}''_B) > \Pi^h_A(\mathbf{x}'_A, \mathbf{x}''_B)$.

Adding up the above inequalities, it follows that

$$\Pi^h_A(\mathbf{x}'_A, \mathbf{x}'_B) + \Pi^h_A(\mathbf{x}''_A, \mathbf{x}''_B) > \Pi^h_A(\mathbf{x}''_A, \mathbf{x}'_B) + \Pi^h_A(\mathbf{x}'_A, \mathbf{x}''_B). \tag{2.14}$$

Repeating the argument for party $B$,

$$\Pi^h_B(\mathbf{x}'_A, \mathbf{x}'_B) + \Pi^h_B(\mathbf{x}''_A, \mathbf{x}''_B) > \Pi^h_B(\mathbf{x}'_A, \mathbf{x}''_B) + \Pi^h_B(\mathbf{x}''_A, \mathbf{x}'_B). \tag{2.15}$$

It is easy to show from (2.14) that

$$v^h_A(\mathbf{x}'_A, \mathbf{x}'_B) + v^h_A(\mathbf{x}''_A, \mathbf{x}''_B) > v^h_A(\mathbf{x}''_A, \mathbf{x}'_B) + v^h_A(\mathbf{x}'_A, \mathbf{x}''_B). \tag{2.16}$$

Multiplying (2.16) by -1, and adding 2 on both sides,

$$(1-v^h_A(\mathbf{x}'_A, \mathbf{x}'_B)) + (1-v^h_A(\mathbf{x}''_A, \mathbf{x}''_B)) < (1-v^h_A(\mathbf{x}''_A, \mathbf{x}'_B)) + (1-v^h_A(\mathbf{x}'_A, \mathbf{x}''_B)). \tag{2.17}$$

Finally, adding $\alpha_B\sigma(\mathbf{x}'_B)$ and $\alpha_B\sigma(\mathbf{x}''_B)$ to both sides of (2.17),

$$\Pi^h_B(\mathbf{x}'_A, \mathbf{x}'_B) + \Pi^h_B(\mathbf{x}''_A, \mathbf{x}''_B) < \Pi^h_B(\mathbf{x}''_A, \mathbf{x}'_B) + \Pi^h_B(\mathbf{x}'_A, \mathbf{x}''_B), \tag{2.18}$$

which stands in contradiction with (2.15). Hence, the equilibrium is unique. ∎

Motivated once again by Alesina and Angeletos (2005a) and Dimick, Rueda and Stegmueller (2017), suppose that the other-regarding utility is smooth and the welfare effect for voter $i$ of a marginal change in his disposable income is invariant to the income of the others.[13] Then, under assumption **A2***, condition $\mathbb{C}^h$ takes a much simpler form, which relates easily to the Lindbeck–Weibull condition.

To elaborate, define for each group $i \in N$, the index $\left(\sum_{j\in N} \xi^h_{ij}(\mathbf{x})\right)^{-1}$, which measures the overall concavity of the utility function $U^h_i(\cdot)$ at $\mathbf{x} \in X$, where $\xi^h_{ij}(\mathbf{x}) = -\frac{[\partial U^h_i(\mathbf{x})/\partial x_j]^2}{\partial^2 U^h_i(\mathbf{x})/\partial x_j \partial x_j}$. Likewise, given a strategy profile $(\mathbf{x}_A, \mathbf{x}_B) \in X \times X$, define the logarithmic rate of change of the probability density $f_i$ as the ratio

[13] Technically, $\frac{\partial^2 \sigma^h(\mathbf{y})}{\partial y_i \partial y_j} = 0$ for all $i \neq j$, $i, j \in N$, and all $\mathbf{y} \in Y$.

$r_i(t_i^h) = \frac{f_i'(t_i^h)}{f_i(t_i^h)}$.[14] It is easy to see that condition $\mathbb{C}^h$ requires that for all $i \in N$,

$$\sup_{t_i^h} r_i(t_i^h) \leq \inf_{\mathbf{x}\in X} \left( \sum_{j\in N} \xi_{ij}^h(\mathbf{x}) \right)^{-1}. \tag{2.19}$$

Notice from (2.19) that if voters are purely selfish, that is, if $\alpha_i = 0$ for all $i \in N$, then condition $\mathbb{C}^h$ reduces simply to Lindbeck and Weibull's (1987) sufficient condition, namely, $\sup r_i(t_i^h) \leq \inf \left(\xi_{ii}^h(\mathbf{x}_C)\right)^{-1}$. The reason is the second-order cross derivatives of the vote shares are all null without other-regarding utility, which simplifies greatly the Hessian matrix of the function $v_C$. By contrast, in the presence of other-regarding concern, the marginal increase in the percentage of votes that one party obtains by changing group $i$'s transfers varies with the transfers allocated to group $j \neq i$, making the cross derivatives nonzero.

## 2.5 Conclusion

We prove that, under fairly general conditions, the modified probabilistic voting model with other-regarding preferences has a unique pure-strategy equilibrium. The proof rests on standard existence results for simultaneous games with a continuum of pure strategies.To guarantee the strict quasi-concavity of the conditional payoff functions, which are continuous in the strategy space, we impose a sufficient condition that bears similarities with Lindbeck and Weibull's (1987). The condition we establish to ensure equilibrium existence demands that the rate at which the percentage of votes of each party varies as result of changes in the relative welfare (utility differential) of the groups be limited by the *overall concavity* of voters' utility function, imposing ipso facto an upper bound on the rate of change of the second term of the party's conditional payoff function. The difficulty of the proof lies in the fact that for inequality aversion preferences, differentiability is not guaranteed and as a result, the negative definiteness of the Hessian matrix only implies the local concavity of parties' payoff functions. To overcome this challenge, we first show that under the proposed condition, the gradients of parties' payoff functions are monotone decreasing in a subset of the

[14]In the uniform case, for instance, this ratio is equal to zero, meaning that changes in the utility differential affect the marginal vote-returns of each party at a constant rate.

domain where differentiability is assumed. However, as such subset is not guaranteed to be convex, we need to show that such condition is sufficient to ensure that the payoff function has a support at each alternative in the policy space.

This restriction is stronger than Lindbeck and Weibull's (1987) condition due to the nonsmoothness of voters' utility functions, which relates in a non-trivial way the margin vote share of each income group with the transfers received by the other groups. However, it is satisfied in a number of meaningful cases, among which we find the uniform distribution case of ideological preferences studied by Lindbeck and Weibull (1987). Also note that although in the model above, we focus our attention on other-regarding preferences, all the results in this paper can be easily extended to other utility representations satisfying Lipschitz continuity.

# Chapter 3

# Core Existence with Nonsmooth Preferences

## 3.1 Introduction

In the spatial theory of voting, alternative policies are seeing as elements in a convex multidimensional space over which a collection of voters have continuous and convex preferences. In this setting, an alternative is preferred over another under a voting rule if there exists a winning coalition supporting such alternative. The set of alternatives that are stable under pairwise voting, that is, it is not strictly preferred by any other alternatives, is said to constitute the core of the voting rule. Since majority rule plays an important role in our decision-making apparatus, the core of the majority rule, so-called majority core, has been received much attention from political scientists and economists alike. Given the observation that the majority-rule decision-making does not always produce a stable outcome, for many years, there has been a huge body of research focusing the conditions under which the majority core is nonempty. Plott's (1967) necessary and sufficient conditions for majority core nonemptiness under the standard differentiability assumption, states that given an odd number of agents, at any majority core alternative, there must be one agent whose gradient equals to zero (median voter in all directions), and suppose that such agent is unique, the gradients of other voters must satisfy radial symmetry, that is, for each agent, there exists a partner whose gradient vector points exactly in the opposite direction. Under such severe symmetric condition, the existence of Condorcet winner is generically unlikely and given a rich enough policy domain, a cycle of pairwise voting can be constructed that winds it way through the entire set of alternatives (McKelvey, 1976; Schofield, 1978).

The current literature on core nonemptiness has focused on the continuous and convex preference domain where each individual's utility function is continuously differentiable everywhere over the policy space. In this paper, we contribute to the literature on spatial voting theory by relaxing the differentiability assumption and considering a broad class of nonsmooth utility functions, under which a stable well-defined majority-rule equilibrium might exist, even in a multidimensional policy space. The nonsmoothness might arise because of the introduction of social preferences which are not guaranteed to be continuously differentiable such as Fehr and Schmidt's (1999) model of inequity aversion; or because voters' preferences over the policy space are represented by Minkowski distance of order 1 or $\infty$. The nonsmooth utility representation does not necessarily violate the convexity of preferences or the pseudo-concavity of utility function. However, as the set of profitable deviations for each agent is now a convex cone instead of a half space, a decisive coalition forms less frequently, which in turn leads to a less restrictive condition for core nonemptiness. More specifically, we assume that each voter in our model has convex preferences, which can be represented by a Lipschitz continuous function. As a result, the derivative of the individual's utility function is not represented by a gradient vector but by a set of *supergradient* vectors, called *superdifferential*. Using the concept of superdifferential, we generalize all the previous characterization results which are based on smooth analysis, adapting it conveniently to accommodate for the nonsmoothness of the utility function. The aim of this paper is not only to show that the core can be nonempty in high-dimensional spaces, but to also characterize the core alternative whenever it exists.

The relaxation of the differentiability assumption in spatial voting model has been also studied by Sloss (1973). She identifies the fact that as the set of profitable deviations forms an open convex cone when the utility function is not differentiable, the likelihood that a stable outcome exists increases. To characterize the sufficient and the necessary conditions for core nonemptiness, she imposes a condition which basically "approximates" the nondifferentiable utility function by a differentiable function and derives a condition similar to that of Plott (1967). In this paper, we argue that such approximation technique excludes situations in which stable outcomes are more likely to exist. To support our claim, we rely on a class of nondifferentiable utility functions that has been extensively studied, called the $\ell_1$ metric or the city block preferences. Agents with city block preferences evaluate each alternative by the weighted sum of the distance on each policy dimension to their ideal point. The problem was intitally studied by Rae and Taylor (1971), Wendell and Thorson (1974) and McKelvey and Wendell (1976). Humphreys and Laver (2010) extends the result by identifying necessary and sufficient conditions for the existence of equilibrium in a more general settings. Their result does not only suggest that the majority core is not generically empty, but it also shows that the set of equilibrium social outcomes is in fact equivalent to the dimension-by-dimension median of the agents' ideal points.

The rest of the paper is organized as follows. In Section 3.2, we describe the spatial framework and provide a simple characterization result for $q$-rules, similar to the case with differentiability assumption. Section 3.3 presents the main result of the paper. In Section 3.4, we provide two examples of nonsmooth utility functions (inequity aversion and city-block preferences) to illustrate the application of our result. Section 3.5 concludes the paper with a summary and remarks.

## 3.2 Preliminaries

Let $N = \{1, \ldots, n\}$ be the set of voters, and let $X \subseteq \mathbb{R}^d$ be a non-empty set of alternatives, which can be modeled as a convex subset of $d$-dimensional Euclidean space. Assume that the preferences of voter $i$ is represented by a *Lipschitz continuous* utility function $u_i : X \to \mathbb{R}$. The preference of voter $i$ is said to be strictly convex if for each $x \in X$, the set of (strictly) profitable deviations

$$\mathcal{D}_i(x) = \{y \in X \,:\, u_i(y) > u_i(x)\} \tag{3.1}$$

is convex. One of the major assumptions in the classical literature on the spatial voting model is that individual preferences are assumed to have continuously differentiable utility representations in which the gradient vectors are well defined over the entire set of alternatives $X$. However, among the models of social preferences, utility differentiability is not always guaranteed. For instance, in the case of inequity aversion model of Fehr and Schmidt (1999), the utility function has a kink and is not differentiable at each individual's reference point. There are other prominent examples of other-regarding behavior relevant for distributive politics that share this nondifferentiability property, including maximin and quasi-maximin preferences, Bolton and Ockenfels' (2000) inequality aversion, etc. Note that nondifferentiability does not contradict the results that is based on ordinal analysis and hence, a more general theorem is needed to characterize the core point in a cardinal environment.

For nonsmooth utility functions, instead of using the gradient vector, we define the notion of *supergradient* and *superdifferential*. For each voter $i \in N$, a vector $p_i(x) \in \mathbb{R}^d$ is a supergradient of voter $i$ evaluated at alternative $x \in X$ if for every alternative $y \in X$, it satisfies the supergradient inequality:

$$u_i(x) + p_i(x) \cdot (y - x) \geq u_i(y). \tag{3.2}$$

For strictly concave function $u_i$, the set of all supergradient of $u_i$ at $x$ is called the superdifferential of $u_i$ at $x$, and is denoted $\partial u_i(x)$. Note that the gradient of a concave function at a point of differentiability is also a supergradient. In fact, the converse is also true, if $\partial u_i(x)$ is a singleton, then $u_i$ is differentiable at $x$. Following Rockafellar (1970),the one-sided directional derivative of $u_i(.)$ in the direction of $r \in \mathbb{R}^d$ defined at $x$, denoted $u_i'(x, r)$ can be related to the superdifferential in the following manner:

$$u_i'(x, r) = \lim_{\delta \to 0} \frac{u_i(x + \delta r) - u_i(x)}{\delta} = \inf_{p_i(x)} p_i(x) \cdot r. \tag{3.3}$$

The utility function of voter $i$ is assumed to be pseudo-concave at $x$ if and only if for every $x, y \in X$.

$$u_i(y) > u_i(x) \implies p_i(x) \cdot (y - x) > 0 \text{ for all } p_i(x) \in \partial u_i(x). \tag{3.4}$$

In words, if individual $i$ weakly prefers $y$ to $x$, then arbitrarily small deviations from $x$ in the direction of $y$ should not make agent $i$ worst off. Suppose $u_i$ represents a strictly convex preference order of agent $i$. Given the convexity of $X$, the pseudo-concavity of $u_i$ insures that there exists an alternative, say $\hat{x}_i \in \mathbb{R}^d$ at which $0 \in \partial u_i(\hat{x}_i)$, which is voter $i$'s most preferred alternative.[1] Let

[1] Note that for a conic restriction, the necessary conditions at the core only consider marginal changes in any direction from a given alternative $x$, and as a consequence, global assumptions such as pseudo concavity of utilities are not used. Then the supergradient inequality (3.2) can be defined by considering a small deviation in the direction of $y$ in the open ball centered at

$u = (u_1, \ldots, u_n) : X \to \mathbb{R}^n$ denote a utility profile, and define $U(X)^n$ to be the space of all Lipschitz continuous utility profiles on $X$.

A (strict) collective preference relation $\succ_{\mathcal{L}}$ over the set $X$ is characterized by a collection $\mathcal{L}$ of decisive coalitions of voters: for all $x, y \in X$, $x \succ_{\mathcal{L}} y$ if and only if there exists $L \in \mathcal{L}$ such that $u_i(x) > u_i(y)$ for all $i \in L$. The collection $\mathcal{L}$ is required to be monotonic: $L \in \mathcal{L}$ and $L \subseteq L'$ imply $L' \in \mathcal{L}$, and proper: $L \in \mathcal{L}$ implies $N \setminus L \notin \mathcal{L}$. A particular class of aggregation rules that attracts our attention are the weighted $q$-rules, that is, the collections of decisive coalitions are of the form

$$\mathcal{L} = \{L \subseteq N : |L| \geq q\} \equiv \mathcal{L}_q. \tag{3.5}$$

where the quota $q$ satisfies $\frac{n}{2} < q \leq n$. Then the core of a $q$-rules, so called $q$-core, is the set of maximal elements in $X$ under the binary relation $\succ_q$ defined by $\mathcal{L}_q$, i.e. it is

$$C(\mathcal{L}_q) = \{x \in X \ : \text{ there does not exist } y \in X \text{ and } L \in \mathcal{L} \\ \text{such that } u_i(y) > u_i(x) \text{ for all } i \in L\}. \tag{3.6}$$

Alternatively, the set of core alternatives can be characterized as the Pareto set of all decisive coalitions, that is: $x \in C(\mathcal{L}_q)$ if and only if, for any decisive coalition $L \in \mathcal{L}_q$, we have $\bigcap_{i \in L} \mathcal{D}_i(x) = \emptyset$.

Define for any agent $i \in N$ the following derived cone at the alternative $x \in X$:

$$C_i^+(x) = \{r \in \mathbb{R}^d : r \cdot p_i(x) > 0 \text{ for all } \ p_i(x) \in \partial u_i(x)\},$$
$$C_i^-(x) = \{r \in \mathbb{R}^d : r \cdot p_i(x) < 0 \text{ for some } \ p_i(x) \in \partial u_i(x)\},$$
$$C_i^0(x) = \{r \in \mathbb{R}^d : \inf_{p_i(x) \in \partial u_i(x)} r \cdot p_i(x) = 0\}.$$

It is trivial that if $0 \in \partial u_i(x)$, then $C_i^+(x) = \emptyset$. Observe that for any alternative $x \in X$ and utility profile $u \in U(X)^N$, if $0 \notin \partial u_i(x)$ then the preference cone $C_i^+(x)$ is an open convex cone in $\mathbb{R}^d$ bounded by the indifference hyperplane $C_i^0(x)$. When the utility function is continuously differentiable, for any $i \in N$ such that $\nabla u_i(x) \neq 0$, the preference cone $C_i^+(x)$ is the half space characterized by $\{r \in \mathbb{R}^d : r \cdot \nabla u_i(x) > 0\}$ and $C_i^0(x)$ is the orthogonal hyperplane of $\nabla u_i(x)$. Following McKelvey and Schofield (1987), given the pseudo-concavity of

the alternative $x$.

the utility profile, we have:

**Lemma 3.1** *Let the utility profile $u \in U(X)^n$ and alternative $x \in X$. Then $x \in C(\mathcal{L}_q)$ if and only if $\bigcap_{i \in L} C_i^+(x) = \emptyset$ for all $L \in \mathcal{L}_q$.*

Given voters' superdifferential profile $\{\partial u_i(x)\}_{i \in N}$, define for any $L \subseteq N$, the semi-positive cone generated by $\{\partial u_i(x)\}_{i \in L}$,

$$\mathcal{P}_L(x) = \Big\{ v \in \mathbb{R}^d : v = \sum_{i \in L} \alpha_i p_i(x), \\ \text{for all}\, p_i(x) \in \partial u_i(x), \alpha_i \geq 0 \ \forall i \in L \text{ and } \exists i \in L, \alpha_i \neq 0 \Big\}. \quad (3.7)$$

We use the convention that $\mathcal{P}_{\emptyset}(x) = \emptyset$. Similar to the result of Smale (1973), the next result trivially extends the characterization theorem of the core to the domain of Lipschitz continuous utility functions.

**Lemma 3.2** *Let the utility profile $u \in U(X)^n$ and the alternative $x \in X$. Then $x \in C(\mathcal{L}_q)$ if and only if $0 \in \bigcap_{L \in \mathcal{L}_q} \mathcal{P}_L(x)$.*

**Proof.** *(Necessity)* Fix a profile $u \in U(X)^n$ and alternative $x \in X$. Let $x \in C(\mathcal{L}_q)$ and suppose, for contradiction, that there exists $L \subseteq \mathcal{L}_q$ such that $0 \notin \mathcal{P}_L(x)$. Then the Separating Hyperplane Theorem implies that there exists $r \in \mathbb{R}^d$ such that, for all $v \in \mathcal{P}_L(x)$, $r \cdot v > 0$. As $\partial u_i(x) \subseteq \mathcal{P}_L(x)$ for all $i \in L$, it implies that $r \cdot p_i(x) > 0$ for all $i \in L$ and $p_i(x) \in \partial u_i(x)$. But since $r \in C_i^+(x)$ for all $i \in L$, this contradicts $x \in C(\mathcal{L}_q)$.

*(Sufficiency)* Suppose that $0 \in \mathcal{P}_L(x)$ for some $x \in X$, that is, $0 = \sum_{i \in L} \alpha_i p_i(x)$ for some $(\alpha_i)_{i \in N}$ that satisfy (3.7) and $p_i(x) \in \partial u_i(x)$. Assume, to the contrary, that there exists some $y \in X$ and $L \in \mathcal{L}_q$ such that $u_i(y) > u_i(x)$ for all $i \in L$. Then by the pseudo-concavity of $u_i$, we have $p_i(x) \cdot (y - x) > 0$ for all $p_i(x) \in \partial u_i(x)$ and $i \in L$. Then $\sum_{i \in N} \alpha_i p_i(x) \cdot (y - x) > 0$ for all $p_i(x) \in \partial u_i(x)$, contradiction. ∎

## 3.3 Main Result

Before stating the result of this section, we gather some notation and definitions.

- Given two non-zero vectors $p, q \in \mathbb{R}^d$, a vector $r \in \mathbb{R}^d$ is a *semi-positive combination* of $p$ and $q$ there exist scalar $\alpha, \beta \geq 0$ (not both zero) such that $r = \alpha p + \beta q$.

- A subset $C \subseteq \mathbb{R}^d$ is a *convex cone* if for all $x, y \in C$, every semi-positive combination of $x$ and $y$ is contained in $C$, and for an arbitrary set $Y \subseteq \mathbb{R}^d$, we can define the *conic hull* of $Y$, denoted cone$Y$, as the intersection of all convex cones $C$ such that $Y \subseteq C$. The *convex hull* of $Y$, denoted conv$Y$, is the intersection of all convex sets containing $Y$.

- A convex cone is *pointed* if $0 \notin C$; it is *finitely generated* if there exist $z^1, \ldots, z^s \in \mathbb{R}^d$ such that $C = \text{cone}\{z^1, \ldots, z^s\}$; and in this case, the set $\{z^1, \ldots, z^s\}$ is a generator of $C$, and the ranks of all generators are equal to the dimension of $C$.

- A set $\{z^1, \ldots, z^s\}$ of vectors is *semi-positively independent* if $0 \notin \text{cone}\{z^1, \ldots, z^s\}$, and otherwise it is *semi-positively dependent.*

- For any set $M \subseteq N$ and any positive integer $\kappa$, the power set of $M$ with cardinality $\kappa$, denoted $\mathcal{P}_\kappa(M)$, is the set of all subsets of $M$ of cardinality equal to $\kappa$, that is, $\mathcal{P}_\kappa(M) = \{M' \subseteq M \text{ such that } |M'| = \kappa\}$.

Given these notations, the formal statement of the main result in Duggan (2018) which will be used latter in the proof of our Theorem 3.2 is as follows:

**Theorem 3.1 (Duggan, 2018)** *Let $x \in intX$, let $K = \{i \in N : p_i(x) = 0\}$ consist of the voters with zero gradient at $x$, let $X \subseteq \mathbb{R}^d$ be a pointed, finitely generated, convex cone with dimension $s$, let*

$$G^+ = \{i \in N : p_i(x) \in C\} \quad \textit{and} \quad G^- = \{i \in N : -p_i(x) \in C\} \tag{3.8}$$

*consist of the voters in $N$ with gradients in $C$ and $-C$, respectively, and let*

$$I = \{i \in N \setminus (G^+ \cup G^- \cup K) : p_i(x) \in \textit{span}C\} \tag{3.9}$$

*consist of the remaining voters in $N$ with non-zero gradients contained in the linear subspace spanned by $C$. If $x$ belongs to the q-core, then*

$$\left\lceil \frac{n - |G^+| - |G^-| + |I| + |K|}{2} \right\rceil + \left\lceil \frac{|I|}{s} \right\rceil + |G^+| \leq q - 1. \tag{3.10}$$

One of the main difficulties when applying the above result to an environment without differentiability assumption is that the set of beneficial deviations for each individual voter is no longer an open half space, thus imposing a lack of structure on the problem. As the superdifferential $\partial u_i(x)$ is a closed convex set, we assume further that for any $i \in N$ and alternative $x \in X \setminus \{\hat{x}_i\}$, there exists a finite set of $k_i$ semi-positively independent vectors $V_i = \{v_i^1, \dots, v_i^{k_i}\} \subseteq \partial u_i(x)$ such that each supergradient vector $p_i(x) \in \partial u_i(x)$ can be written as a semi-positive combination of $v_i^1, \dots, v_i^{k_i}$. That is, for every $p_i(x) \in \partial u_i(x)$, we have

$$p_i(x) = \alpha_i^1 v_i^1 + \cdots + \alpha_i^{k_i} v_i^{k_i}, \tag{3.11}$$

for some $\alpha_i = (\alpha_i^1, \dots, \alpha_i^{k_i}) \in \mathbb{R}_+^{k_i}$. Note that when $u_i$ is continuously differentiable at $x$, then $k_i = 1$. This assumption holds true for the case of inequity aversion and city-block spatial preferences but fails to accommodate for Sloss's (1973) notion of epsilon-core. Given the pseudo-concavity of $u_i$, it results directly from supergradient inequality that for any alternative $y \in X$,

$$u_i(y) > u_i(x) \implies v_i^k \cdot (y - x) > 0 \quad \text{for all } v_i^k \in V_i. \tag{3.12}$$

As Duggan's (2018) result applies for an arbitrary number of voters, we can characterize the necessary restriction at the core without differentiability by considering the following auxiliary spatial voting model: a society $\tilde{N}$ of $\sum_{i \in N \setminus K} k_i$ singular voters (or 'self');[2] each self $i^k$ is endowed with a smooth utility function the gradient vector at $x$ of which is given by $v_i^k \in Z_i$; and assign a positive weight $\alpha_i^k$ to each self $i^k$ such that $\alpha_i = (\alpha_i^1, \dots, \alpha_i^{k_i}) \in \mathbb{R}_+^{k_i}$ and $\sum_{k=1}^{k_i} \alpha_i^k = 1$. With a slight abuse of notation, denote $\alpha(M) = \sum_{i^k \in M} \alpha_i^k$ for any $M \subseteq \tilde{N}$. Note that we only consider selves of agents who do not have their ideal alternatives at $x$, that is, $v_i^k \neq 0$ for all $i^k \in \tilde{N}$. Under these assumptions, the number of agents who consider a deviation in the direction $r \in \mathbb{R}^d$ beneficial can be written as follows:

$$|\{i \in N : u_i'(x, r) > 0\}| = \min_{\{\alpha_i\}_{i \in N}} \alpha(\{i^k \in \tilde{N} : v_i^k \cdot r > 0\}). \tag{3.13}$$

Notice that an alternative in the $q$-core of the auxiliary spatial voting model for

[2]There are literature on the theory of multiple self, proposing the idea that the individual person may be seen as a set of sub-inidividuals, each with different interest, moral principle and belief. See Elster (1987) for some applications of the multiple self theory in philosophy, psychology and economics.

any $\alpha = \{\alpha_i\}_{i\in N}$ will also be in the $q$-core for the spatial voting model. However, the converse does not hold as a deviation from the alternative in consideration is beneficial for an agent $i$ if and only if all of his selves $i^k$ agree on such deviation. Given the foregoing observation, if $x$ belongs to the $q$-core, then it must be that (i) no coalition of selves has its sum of weights larger than the quota; and (ii) the weight of a self is counted toward the coalition only if all other selves of him is also included in the coalition, as implied by the following theorem.

**Theorem 3.2** *Let $x \in intX$, let $C \subset \mathbb{R}^d$ be a pointed, finitely generated, convex cone with dimension $s$, let*

$$\tilde{G}^+ = \{i^k \in \tilde{N} : v_i^k \in C\} \quad \textit{and} \quad \tilde{G}^- = \{i^k \in \tilde{N} : -v_i^k \in C\} \tag{3.14}$$

*consist of the selves in $\tilde{N}$ with gradients in $C$ and $-C$, respectively; let*

$$\tilde{I} = \{i^k \in \tilde{N} \setminus (\tilde{G}^+ \cup \tilde{G}^-) : v_i^k \in spanC\} \tag{3.15}$$

*consist of the selves in $\tilde{N}$ with gradients contained in the linear subspace spanned by $C$; and let $\tilde{T} = \tilde{N} \setminus (\tilde{G}^+ \cup \tilde{G}^- \cup \tilde{I})$ be the remaining selves. If $x$ belongs to the $q$-core then*

$$\min_{\{\alpha_i\}_{i\in N}} \min_{\tilde{T}' \in \mathcal{P}_{\lceil \frac{|\tilde{T}|}{2} \rceil}(\tilde{T})} \min_{\tilde{I}' \in \mathcal{P}_{\lceil \frac{|\tilde{I}|}{s} \rceil}(\tilde{I})} \alpha(\tilde{G}^+) + \alpha(\tilde{I}') + \alpha(\tilde{T}') < q \tag{3.16}$$

**Proof of Theorem 3.2** Assume that $x$ is an interior $q$-core aternative and $C$ is a pointed, finitely generated, convex cone with dimension $s$. Suppose, for contradiction that

$$\min_{\{\alpha_i\}_{i\in N}} \min_{T' \in \mathcal{P}_{\lceil \frac{|T|}{2} \rceil}(T)} \min_{I' \in \mathcal{P}_{\lceil \frac{|I|}{s} \rceil}(I)} \alpha(G^+) + \alpha(I') + \alpha(T') \geq q. \tag{3.17}$$

Let $\{z^\ell \in \mathbb{R}^d : \ell \in L\}$ be a finite, linear independent set, indexed by $L = \{1, \ldots, s\}$ that generates $C$. Similar to Duggan (2018), we apply the duality theorem to the cone $C$. Given any vector $p \in \mathbb{R}^d$, let $H_p^- = \{v \in \mathbb{R}^d : p \cdot v \leq 0\}$ be the closed halfspace in $\mathbb{R}^d$ generated by $p$. Then the dual of $C$ is

$$C^* = \{v \in \mathbb{R}^d : \text{for all}\, \ell \in L,\, z^\ell \cdot v \leq 0\} = \bigcap_{\ell \in L} H_{z^\ell}^-, \tag{3.18}$$

and given $C$ is of dimension $s$, the dual cone $C^*$ is also generated by a set of $s$ extreme vectors. In particular, denote the set of $s$ extreme vectors of $C^*$ by $\{w^\ell \in \mathbb{R}^d : \ell \in L\}$ such that for all $\ell \in L$ and $k \in L \setminus \{\ell\}$, we have $w^\ell \cdot z^k = 0$. By duality, the dual of $C^*$ is $C$, so that

$$C = \{v \in \mathbb{R}^d : \text{for all}\, \ell \in L,\, w^\ell \cdot v \leq 0\} = \bigcap_{\ell \in L} H^-_{w^\ell}. \tag{3.19}$$

For each $\ell \in L$, let $\tilde{G}^\ell = \{i^k \in \tilde{N} : v_i^k \cdot w^\ell < 0\}$ consist of the selves of $\tilde{I}$ whose gradients have negative dot product with $w^\ell$. Note that for each self $i^k \in \tilde{I}$, we have $v_i^k \notin -C = -\bigcap_{\ell \in L} H^-_{w^\ell}$, thus, there exists $\ell_i^k \in L$ such that $v_i^k \cdot w^{\ell_i^k} < 0$, which implies $i^k \in \tilde{G}^{\ell_i^k}$. Therefore, $\tilde{I} = \bigcup_{\ell \in L} \tilde{G}^\ell$ and it follows that there is at least one $\ell^* \in L$, such that

$$|\tilde{G}^{\ell^*}| \geq \left\lceil \frac{|\tilde{I}|}{s} \right\rceil. \tag{3.20}$$

Note that for all $i^k \in \tilde{G}^+$, we have $v_i^k \in C \subseteq H^-_{w^{\ell^*}}$, which implies $v_i^k \cdot w^{\ell^*} \leq 0$. Indeed, since $C$ is pointed, it follows that $0 \notin C$ and by the separating hyperplane theorem, there exists a vector $y \in \mathbb{R}^d$ such that for all $i^k \in \tilde{G}^+$, we have $v_i^k \cdot y > 0$. Let $\epsilon > 0$ and $w^\epsilon = w^{\ell^*} - \epsilon y$, we then have

$$v_i^k \cdot w^\epsilon = v_i^k \cdot w^{\ell^*} - \epsilon z_i^k \cdot y < 0 \tag{3.21}$$

for all $i^k \in \tilde{G}^+$, and we can choose $\epsilon > 0$ sufficiently small that $v_i^k \cdot w^\epsilon < 0$ for all $i^k \in \tilde{I}$.

If $d = s$, then we define $r = -w^\epsilon$. Otherwise, if $d > s$, there exists a vector $r^*$ that is orthogonal to span$C$ and such that for all $i^k \in \tilde{T}$, $v_i^k \cdot r^t \neq 0$. Furthermore, reversing the direction of $r^m$ if needed, we can assume without loss of generality that

$$|\{i^k \in \tilde{T} : v_i^k \cdot r^m > 0\}| \geq \frac{|\tilde{T}|}{2}. \tag{3.22}$$

Finally, we define $r = r^* - \delta w^\epsilon$, where $\delta > 0$ is chosen sufficiently small so that for all $i^k \in \tilde{T}$, $v_i^k \cdot r \neq 0$ and has the same sign as $v_i^k \cdot r^*$. By construction, for all $i^k \in \tilde{G}^+ \cup \tilde{G}^{\ell^*}$, $r^*$ is orthogonal to all the self gradients, which implies

$$v_i^k \cdot r = v_i^k \cdot r^* - v_i^k \cdot (\delta w^\epsilon) = -\delta v_i^k \cdot w^\epsilon > 0. \tag{3.23}$$

Combining (3.22) and (3.23), we have

$$
\begin{aligned}
&|\{i \in N : u_i'(x,r) > 0\}| \\
&\qquad = \min_{\{\alpha_i\}_{i\in N}} \alpha(i^k \in \tilde{N} : v_i^k \cdot r > 0) \\
&\qquad = \min_{\{\alpha_i\}_{i\in N}} \alpha(\tilde{G}^+) + \alpha(\{i^k \in \tilde{I} : v_i^k \cdot r^m > 0\}) + \alpha(\{i^k \in \tilde{T} : v_i^k \cdot r^m > 0\}) \\
&\qquad \geq \min_{\{\alpha_i\}_{i\in N}} \min_{\tilde{T}' \in \mathcal{P}_{\lceil \frac{|\tilde{T}|}{2} \rceil}(\tilde{T})} \min_{\tilde{I}' \in \mathcal{P}_{\lceil \frac{|\tilde{I}|}{s} \rceil}(\tilde{I})} \alpha(\tilde{G}^+) + \alpha(\tilde{I}') + \alpha(\tilde{T}') \\
&\qquad \geq q,
\end{aligned}
$$

where the last inequality follows trivially from (3.17). However, since $x$ belongs to the interior of $X$, we can choose $\gamma > 0$ sufficiently small such that $x + \gamma r \in X$, and for each voter $i$, $u_i'(x,r) > 0$, then $u_i(x + \gamma r) > u_i(x)$. But then the latter inequality holds for at least $q$ voters, contradicting the assumption that $x$ belongs to the $q$-core. ∎

## 3.4 Examples

This section provides two simple applications of Theorem 3.2 for the case of majority core with agents exhibiting inequity aversion and city-block preferences.

### 3.4.1 Inequity Aversion

We consider the canonical distribution problem where voters admit social preferences modeled after (reference-dependent) inequity aversion of Fehr and Schmidt (1999). Consider the case with three agents $N = \{1, 2, 3\}$. Denote the income of each agent $i \in N$ by $x_i > 0$. Let $\sum_{i \in N} x_i = 1$ be the budget constraint of the economy and denote the set of all possible distribution of wealth by $X = \{x = (x_1, x_2, x_3) : \sum_{i\in N} x_i = 1\} \subset \mathbb{R}^3$.

The utility function of each voter $i \in N$ associated with each income distribution $x \in X$ is given by

$$
u_i(x) = x_i - \left[ \sum_{j \neq i} \gamma_{ij} \max(x_j - x_i; 0) + \sum_{j \neq i} \beta_{ij} \max(x_i - x_j; 0) \right], \tag{3.24}
$$

where the first expression $x_i$ is the linear self-regarding utility function and the second expression represents voter $i$'s inequity aversion term. The parameters $\beta_{ij}, \gamma_{ij}$ is the agent $i$'s altruism and envy weight, respectively, towards agent $j$, where $\beta_{ij} \leq \gamma_{ij}$ for all $i, j \in N$. The intuition is that individuals are more selfish than altruistic, and consequently that they are more concerned with disadvantageous inequity. It is trivial to check that the above utility function satisfies Lipschitz continuity but it is not continuously differentiable.

In a distribution problem where agents are purely selfish, majority induced social preference is intransitive and the majority core is generically empty. However, as agents admit inequity averse preferences, the nonsmooth of the utility function can result in the nonemptiness of the majority core. Consider the egalitarian distribution $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Suppose that the agent are partially selfish, that is, for each $i \in N$, agent $i$ strictly prefer the alternative that distribute $i$ with everything to any other alternatives: for any $j, k \in N \setminus \{i\}$,[3]

$$\begin{cases} 1 - 2\beta_{ij} - \beta_{ik} > 0; \\ 1 - 2\beta_{ik} - \beta_{ij} > 0. \end{cases} \tag{3.25}$$

By definition, $x^*$ is in the majority core if and only if there is no other alternative in $X$ that can beat $x^*$ in pairwise majority voting. Consider an alternative $x' \in X$ and $x' \neq x^*$. Then there exists $i \in N$ such that $x'_i \neq x^*_i$. Without loss of generality, let $i = 1$ and $x'_i > x^*_i$. Since $x' \in X$, then there must exists $j \in N \setminus \{i\}$ such that $x'_j < x^*_j$. Without loss of generality, let $j = 3$. Then, $x^*$ is in the majority core only if

$$\begin{aligned} &\text{either } \frac{\beta_{13} - \beta_{12}}{1 - \beta_{12} - 2\beta_{13}} \geq \frac{1 + \gamma_{21} - 2\beta_{23}}{\gamma_{21} + \beta_{23}}; \\ &\text{or } \beta_{23} \geq \frac{1}{3}, \end{aligned} \tag{3.26}$$

and also,

$$\begin{aligned} &\text{either } \frac{\beta_{23} - \beta_{21}}{1 - \beta_{21} - 2\beta_{23}} \geq \frac{1 + \gamma_{12} - 2\beta_{13}}{\beta_{13} + \gamma_{12}}; \\ &\text{or } \beta_{13} \geq \frac{1}{3}. \end{aligned} \tag{3.27}$$

[3]This restriction basically guarantees that the marginal utility of one's own income is strictly positive and the egalitarian distribution is not ideal for any agents, that is, $\hat{x}_i \neq x^*$ for all $i \in N$.

As a result, under conditions (3.26) and (3.27), agent 1 does not have incentives to form a coalition with agent 2 and deviate from $x^*$. Repeating the same argument for the other coalition between agent 1 and 3 and agent 2 and 3, we can construct a set of constraints on $\{\beta_{ij}, \gamma_{ij}\}_{\substack{i \in N \\ j \neq i}}$ which guarantee the nonemptiness of the majority core. The intuition is that agent 1 wants to form a coalition with agent 2 to deviate from $x^*$ by hurting agent 3. However, as agent 2 exhibits significant altruism toward agent 3, no proposals suggested by agent 1 can strictly benefit agent 2. Similarly, agent 2 finds it beneficial to collude with agent 3 but no coalition is formed as agent 3 is highly altruistic toward agent 1. Finally, agent 1's altruism toward agent 2 will also block the coalition between agent 1 and agent 3. Consider the following set of parameters which guarantees $x^*$ belongs to the majority core:

$$\begin{cases} \beta_{13} = \beta_{21} = \beta_{32} = 0.4; \\ \beta_{12} = \beta_{23} = \beta_{31} = 0.1; \\ \gamma_{13} = \gamma_{21} = \gamma_{32} = 0.5; \\ \gamma_{12} = \gamma_{23} = \gamma_{31} = 0.2. \end{cases} \tag{3.28}$$

Given the values of $\{\beta_{ij}, \gamma_{ij}\}_{\substack{i \in N \\ j \neq i}}$, the population of selves $\tilde{N}$ is given by $\{i^k\}_{\substack{i \in N \\ k=1,2}}$ and we can write the explicit coordinates of the self gradient vectors $z_i^k$ for each agent $i \in N$, as follows,

$$\begin{aligned} v_1^1 &= (0.1, -0.3) \quad ; \quad v_1^2 = (1.9, 0.6); \\ v_2^1 &= (0.3, 0.4) \quad ; \quad v_2^2 = (-0.6, 1.3); \\ v_3^1 &= (-0.4, -0.1) \quad ; \quad v_3^2 = (-1.3, -1.9). \end{aligned} \tag{3.29}$$

Now consider $C = \text{cone}\{v_1^2, v_2^2\}$ a pointed (closed) convex cone generated by the gradient vectors of self $v_1^2$ and $v_2^2$. Clearly, we have $G^+ = \{1^2, 2^1, 2^2\}$, and the negative cone $-C$ contains the gradient vectors of exactly two selves $v_3^2$ and $v_1^1$, giving us $G^- = \{3^2, 1^1\}$. Given the dimension of $C$ is 2, equal to the dimension of $X$, we have $T = \emptyset$ and $I = \{3^1\}$. Applying Theorem 3.2 for the necessary restriction at a majority core, we have

$$\begin{aligned} &\min_{\{\alpha_i\}_{i \in N}} \min_{I' \in \mathcal{P}_{\lceil \frac{|I|}{s} \rceil}(I)} \alpha(G^+) + \alpha(I') \\ &= \min_{\{\alpha_i\}_{i \in N} \in [0,1]^6} (\alpha_2^1 + 1 - \alpha_2^1 + \alpha_1^2) + \alpha_3^1 = 1 < q = 2. \end{aligned} \tag{3.30}$$

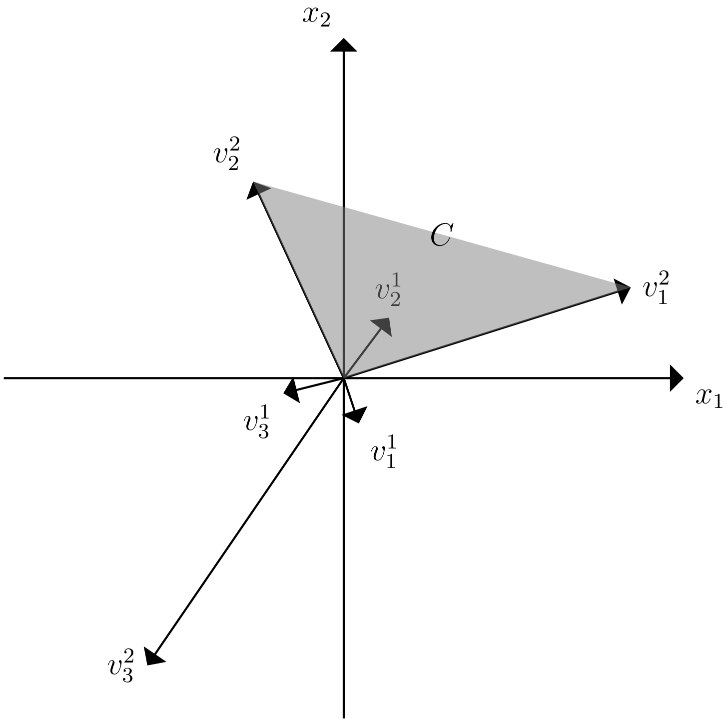


Figure 3.1: The self vectors at egalitarian outcome of agents with inequity aversion.

For the purpose of demonstration, we consider a pointed convex cone $C$ of dimension 1, for example, $C = \text{cone}\{v_1^1\}$. Then we have $G^+ = \{1^1\}$, $G^- = I = \emptyset$, and $T = \tilde{N} \setminus \{1^1\}$. Again, applying Theorem 3.2, we have

$$\begin{aligned} &\min_{\{\alpha_i\}_{i\in N}} \min_{T' \in \mathcal{P}_3(T)} \alpha(G^+) + \alpha(T') \\ = &\min_{\{\alpha_i\}_{i\in N} \in [0,1]^6} \alpha_1^1 + (\alpha_3^1 + \alpha_2^1 + 1 - \alpha_1^1) = 1 < q = 2. \end{aligned} \tag{3.31}$$

### 3.4.2 City-block Preferences

An alternative assumption to (smooth) Euclidean preferences is city-block preferences, which are representable by utility functions that are decreasing in the $\ell_1$-metric distance.[4] Generally, an agent $i \in N$ with ideal point $\hat{x}_i \in \mathbb{R}^d$ admits a $\ell_1$-norm utility representation if there exists a vector of positive scalar $(\beta_m)_{m=1,\dots,d}$

[4]See Eguia (2011, 2013) among others for axiomatic foundation and empirical justification of Minkowski's metric utility function in general and city-block preferences in particular.

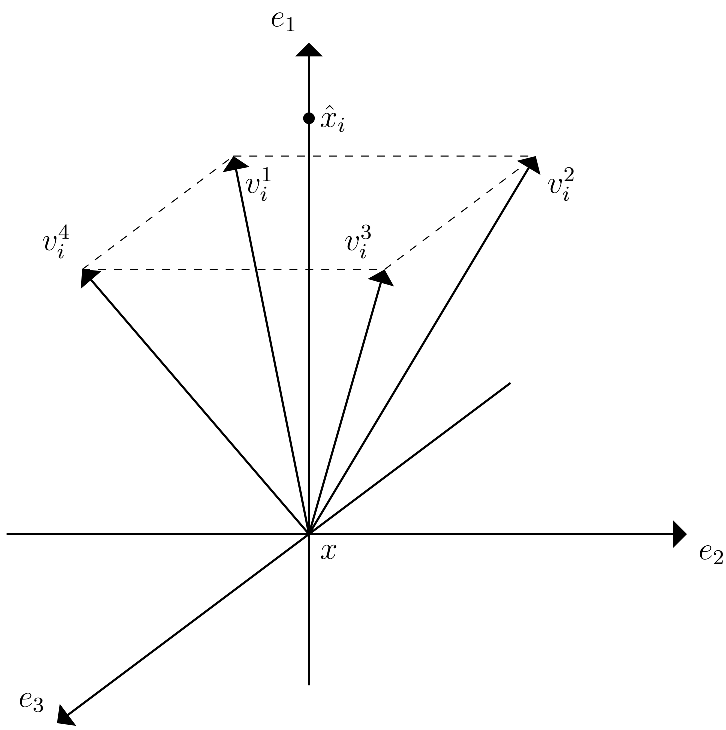

Figure 3.2: The self vectors of agents $i$ with city-block preferences.

such that for any alternative $x \in X$, the utility of agent $i$ is given by

$$u_i(x) = -\sum_{l=1}^{d} \beta_m |x_l - \hat{x}_{im}|, \tag{3.32}$$

where $x_m$ and $\hat{x}_{im}$ is the projected position of $x$ and $\hat{x}_i$ on issue $m$, respectively. In words, agents with city-block preferences calculate the distance between between two alternatives in $\mathbb{R}^d$ by adding up the distance weighted dimension by dimension and they prefer alternatives closer to their ideal alternative according to this notion of distance. The weighting parameters $(\beta_m)_{m=1,\ldots,d}$ reflect the relative view of voters on different political issues. It is common in the literature to assume that all agents attach the same relative weights to the different issues.

Under these assumptions, Humphreys and Laver (2010) derive the the necessary and sufficient restrictions at the core in the following manner: denote $\mathcal{O}$ the collection of $2^d$ open orthant of $\mathbb{R}^d$, each orthant $O \in \mathcal{O}$ can be characterized by a unique vector $w(O) \in W \equiv \{-1, 1\}^d$ such that for any vector $v \in O$, we have $w(O) \cdot v > 0$. Then a "rugged half space" is defined as the closure of a collection, C, of $2^{d-1}$ (open) orthants having the property that there exists a vector $x \in \mathbb{R}^d$

such that for every orthant $O \in C$, we have $w(O) \cdot x > 0$.

**Proposition 3.1 (Humphreys and Laver, 2010)** *Assume a city block preference profile where voters put the same (relative) weight on each issue. Then* $x \in C(\mathcal{L}_q)$ *if and only if every rugged half space contains at least* $(n-q+1)$ *ideal points.*

Now we show that the necessary result of Humphreys and Laver (2010) can be obtained as a straightforward application of Theorem 3.2. It is trivial to check that the $\ell_1$-metric utility function satisfies Lipschitz continuity, that is, it is continuously differentiable almost everywhere in $\mathbb{R}^d$ but a set of alternatives $x$ such that $x_m = \hat{x}_{im}$ for some issue $m = 1, \ldots, d$. Denote $e_m$ the vector forming the $m$th row of the $d \times d$ identity matrix. Then the superdifferentials of an agent endowed with city-block preferences at an alternative $x \neq \hat{x}_i$ is given by

$$\partial u_i(x) = \text{conv}\big\{\{[\text{sign}\,(\hat{x}_{im} - x_m)]\beta_m e_m\}_{m \notin M_i(x)}, \{\pm\beta_m e_m\}_{m \in M_i(x)}\big\}. \tag{3.33}$$

with

$$M_i(x) = \{l \,|\, \hat{x}_{im} = x_m\}. \tag{3.34}$$

Figure 3.2 depicts an agent $i$ with city-block preferences who has his ideal alternative at $\hat{x}_i$ and an alternative $x \in X$ such that $\hat{x}_{i2} = x_2$ and $\hat{x}_{i3} = x_3$. Then the superdifferential of agent $i$ at $x$ can be written by $\partial u_i(x) = \text{conv}\{\beta_1 e_1, \pm\beta_2 e_2, \pm\beta_3 e_3\}$. Consider any pointed, finitely generated convex cone $\bar{C}$ that belongs to the complement of a rugged halfspace $C$ as defined above. Given the superdifferentials of $\ell_1$-metric utility function and the dimension of $\bar{C}$, we have $T = I = \emptyset$. Applying Theorem 3.2, we have

$$\min_{\{\alpha_i\}_{i \in N}} \alpha(G^+) = \#\{i \in N : \hat{x}_i \in \bar{C}\} < q, \tag{3.35}$$

which basically requires that the rugged half space $C$ must contains at least $(n-q+1)$ ideal points. Another form of Minkowski distance that admits a non-smooth utility representation is the $\ell_\infty$-metric or Tchebycheff distance. Tchebycheff distance utility function is mostly used to describe the preferences of agents in a minimax facility allocation problem. Formally, under $\ell_\infty$-norm, the utility function of an agent with ideal point $\hat{x}_i$ is given by

$$u_i(x) = -\max_{m=1,\ldots,d} |x_m - \hat{x}_{im}|. \tag{3.36}$$

Again, the differentials of Tchebycheff distance utility function at an alternative $x \neq \hat{x}_i$ can be written as

$$\partial u_i(x) = \text{conv}\big\{[\text{sign}(\hat{x}_{im} - x_m)]e_m \,|\, m \in M_i(x)\big\}, \tag{3.37}$$

with

$$M_i(x) = \{m \,|\, |\hat{x}_{im} - x_m| = u_i(x)\}. \tag{3.38}$$

Applying similar arguments as above, we can obtain the characterization results for the core under $\ell_\infty$-norm metric.

**Proposition 3.2** *Assume a Tchebycheff distance preference profile. Then* $x \in C(\mathcal{L}_q)$ *if and only if every closed half space through* $x$ *with normal* $w(O) \in W$ *contains at least* $(n - q + 1)$ *ideal points.*

## 3.5 Conclusion

Spatial models of simple majority voting suggest that stable decisions are not likely to exist under general environment. This negative result comes from the fact that for a majority core to exist, it must be some individuals' most preferred alternatives and all other agents' gradient vectors must satisfy the restrictive condition of radial symmetry. In this paper, we briefly show that such difficulties can be overcomed if we allow individuals' utility functions to be Lipschitz continuous instead of continuously differentiable. To deal with the difficulty caused by the absence of differentiable utility functions, we regard each voter's preference at each kinked point as "unanimous" social preferences of a collective of finite "self", each endowed with a smooth preference over the set of alternatives. By applying the "smooth" analysis to the auxiliary society of selves, we generalize the known gradient restriction for the core of an arbitrary quota rule with any utility profile that satifies the general Lipschitz continuity. We also consider two example of nonsmooth utility representation (inequity aversion and city-block preferences) in the context of simple allocation problem to illustrate the application of our results.

However, the paper is incomplete as we also need to identify the critical dimension of the policy space above which the core is generically empty. The previous results by Schofield (1978), McKelvey and Schofield (1987) and Banks (1995) on the critical dimensions for core emptiness are derived by showing the linear dependency on certain subsets of gradient vectorsand then revoking known results from differential topology pertaining to singularities of smooth mappings. As we have obtain similar result on the linear dependency of supergradient vectors (Lemma 3.1), a critical dimension for core emptiness with nonsmooth utility function can be obtained.

# Chapter 4

# Optimal Voting-by-Committees Rules

## 4.1 Introduction

The standard framework of social choice theory assumes that individual preferences over the set of alternatives are binary relations that satisfy completeness, reflexivity, and transitivity. Although completeness is a convenient axiom, it imposes a considerable computational and informational burden to the agents and to the social planner. One main challenge in relaxing preference completeness is that, as the notion of incentive compatibility becomes weaker, the existing results on strategy-proof social choice rules are not necessarily valid. The main purpose of this paper is to elicit the structure of the voting rules that are robust to strategic behavior and "optimal" on the domain of incomplete and separable preferences.

To be more precise, we consider the social choice problem of Barberà, Sonnenschein and Zhou (1991), where society choose a subset from a finite set of indivisible objects (or alternatives). Each member of society is endowed with an asymmetric preference relation over the subsets of alternatives. Like in Barberà et al. (1991), individual preferences are separable, meaning that the consumption of every object affects each agent's welfare separately from the consumption of any other object. However, in a clear departure from previous work, we study separable preference domains where incompleteness (partial orderings) is permitted. This forces us to redefine the notion of incentive compatibility, for which we introduce the concept of *justifiable strategy-proofness*.

Formally, a *voting rule* (mapping each preference profile to an alternative) is justifiable strategy-proof if and only if for each profile of individual preferences and every agent, each possible individual deviation from truth-telling results in a social choice that is strictly worse for the agent in question (bear in mind that preferences are asymmetric). Equipped with this notion of incentive compatibility, in Section 4.3 we show that over the incomplete and separable preference domain, a voting rule is justifiable strategy-proof if and only if it satisfies *monotonicity* and *independence*.[1] This result is used in Theorem 4.1 to show that the set of voting rules that satisfy justifiable strategy-proofness over our preference domain coincides with the family introduced by Barberà et al. (1991), namely, the *voting-by-committees rules*.

Given that the family of voting-by-committees rules is very large, ranging from the dictatorial to the unanimous voting rule, one might pose the following question. Within the family of voting rules that satisfy our notion of incentive compatibility, which rules are most likely to yield an "optimal" social choice? To answer this question, we first study the implications of Pareto efficiency in an incomplete preference domain. As is shown in Proposition 4.3, efficiency turns out to be a weak property. To be precise, it is equivalent to unanimity. The result is encouraging in the sense that, different from Barberà et al. (1991), the set of efficient voting-by-committees rules expands and contains more than just dictatorial rules.[2] However, it also means that the notion of Pareto efficiency is too weak to narrow our search in a meaningful manner.

Since ordinal preferences do not provide any measure of desirability of the different social choices, we turn to a cardinal approach and we study the optimality problem from a utilitarian viewpoint. In our model, individual preferences are not necessarily complete. Therefore, we follow Ok (2002) and adopt a multi-utility representation of preferences, where each partial ordering is represented by a vector-valued utility function. Danan, Gajdos, and Tallon (2015) show that under a multi-utility representation, Harsanyi's aggregation theorem remains valid, and the social welfare function can be written as a set of utilitarian aggregations of individual utility functions. Finally, we compare among the different sets of social welfare functions by invoking several optimal criteria, including *Hurwicz*

[1] For brevity, the reader is referred to Section 4.2.2 for a formal definition of these properties, which were first proposed by Kasher and Rubinstein (1997).

[2] Barberà et al. (1991) show that in a complete and separable preference domain, there exists no efficient, strategy-proof, and nondictatorial voting rule.

*optimism-pessimism*, *minimax regret*, and *distortion minimization*. Our results suggest that the optimal voting-by-committees rules can be characterized as a subfamily of weighted majority rules, with the individual weights representing the level of "caution" or "justice" that the social planner is willing to achieved.

With regard to the literature most closely related to this article, the social choice problem considered here has been studied before by Barberà et al. (1991), Ju (2005), and Nehring and Puppe (2007), among others. These papers have attempted to overcome the Gibbard-Satterthwaite Theorem by characterizing a great variety of strategy-proof social choice rules on restricted preference domains. A recent paper by Hatsumi et al. (2014) suggests that separability constitutes the maximal domain of complete preferences over which a nontrivial rule satisfying strategy-proofness and no-vetoer can exist. One common assumption in this body of research is that the preference relations are represented by complete linear orderings. In this paper, we explore the possibility of extending some of these results to the domain of incomplete and separable preferences.

The literature on the optimality of the social choice rules has been built around Rae's (1969) idea of comparing different voting rules in terms of ex-ante expected utility. Notably, Azrieli and Kim (2014) identify qualified majority rules as interim Pareto efficient within the class of Bayesian incentive compatible social choice rules. Gershkov, Moldovanu and Shi (2017) extend the analysis to the setting with an arbitrary number of agents and alternatives where the privately informed agents have single-crossing and single-peaked preferences. They show that the ex-ante welfare maximizing mechanism in such context falls into the class of sequential voting schemes with flexible majority thresholds. On the other hand, research has been done in the field of computational social choice to relax the information requirement of eliciting full preferences. More specifically, when the communication between the social planner and voters is limited, Boutilier et al. (2015) and Caragiannis and Procaccia (2011) suggest the use of distortion and minimax regret as alternative measurements of a voting rule's performance. However, as strategic voting is not considered in their framework, the family of scoring rules that emerges as optimal in their analysis are subject to strategic manipulation.

The rest of the paper is organized as follows. In Section 4.2, we describe the social choice model and define some important strategic and non-strategic axioms for our analysis. In Section 4.3, we state our characterization results in relation with the axioms defined in Subsection 4.2.2. Section 4.4 focus on the notion of Pareto efficiency in the context of incomplete preferences and it studies its implications. Finally, in Section 4.5, we assume a multi-utility representation for the incomplete preferences and proceed to identify the optimal voting-by-committees rules under different optimal criteria. Section 4.6 concludes the paper with a summary and some remarks.

## 4.2 The Model

### 4.2.1 Basic concepts

We consider a society of $n$ agents who choose collectively a subset from a finite set of $k$ indivisible objects. Let $N$ be the set of agents and $K$ be the set of objects. We assume that $n,\ k \geq 2$. Any subset of $K$ is an (social) alternative and the power set of $K$, $2^K$, describes the set of all possible alternatives. Each agent $i \in N$ is characterized by a preference relation $P_i$ , an asymmetric and transitive partial ordering over the set of all social alternatives.

We focus on the preference domain with the following restriction. A preference relation $P_i$ is separable if for all $x \in K$ and all $X \subseteq K \setminus \{x\}$,

$$[X \cup \{x\}]\, P_i\, X \iff \{x\}\, P_i\, \emptyset. \tag{4.1}$$

Let $\mathcal{S}$ be the domain of all admissible separable preferences. Intuitively, for separable preferences, objects can be partitioned into two categories depending on its ranking compared to the empty set. An object $x \in K$ is considered desirable according to $P_i \in \mathcal{S}$ if for all $X \subseteq K \setminus \{x\}$, $[X \cup \{x\}]\ P_i\ X$. Similarly, object $x$ is undesirable for $P_i$ if for all $X \subseteq K \setminus \{x\}$, $X\ P_i\ [X \cup \{x\}]$. Let $D(P_i)$ be the set of desirable objects according to $P_i$, then the complement of $D(P_i)$ in $K$ is the set of undesirable objects. When there is no confusion, we write $D_i, D_i'$ for $D(P_i), D(P_i')$.

A social choice problem is then represented by a preference profile $P \equiv (P_i)_{i \in N} \in \mathcal{S}^N$. A social choice rule is a function $f : \mathcal{S} \to 2^K$ mapping each preference profile $P \in \mathcal{S}^N$ into a single alternative $X \in 2^K$. We focus on the rules that depend only on the simple information of preferences in terms of desirable and undesirable objects. Formally, a social choice rule is a voting rule if for all $P, P' \in \mathcal{S}^N$, $D(P_i) = D(P_i')$ for all $i \in N$ implies that $f(P) = f(P')$.

Henceforth, we use the following notation. For each $P \in \mathcal{S}^N$ and $x \in K$, let $N_x(P) \equiv \{i \in N : x \in D(P_i)\}$ the set of agents who prefer objects $x$ to be included in the social alternative. For each $P \in \mathcal{S}^N$, each $i \in N$ and $P_i' \in \mathcal{S}$, let $(P_i', P_{-i})$ be the preference profile where agent $i$ has preference relation $P_i'$ and for each agent $j \in N \setminus \{i\}$, agent $j$'s preference is presented by $P_j$.

Suppose that agents' preferences over the set of alternatives $2^K$ are separable but not necessarily complete, in the sense that some pairs of desirable objects are incomparable and some pairs of undesirable objects are incomparable. To be more precise, endow each agent $i \in N$ with a set $\mathcal{S}(P_i) \subset \mathcal{S}$ of *complete* and separable preference relations (also termed rationales or "justification" by Heller (2012)) that share the same set of desirable and undesirable objects. We denote $\tilde{P}_i$ a generic elements of $\mathcal{S}(P_i)$, then $D(\tilde{P}_i) = D_i$ for all $\tilde{P}_i \in \mathcal{S}(P_i)$. Let $\mathcal{S}(P) = \prod_{i \in N} \mathcal{S}(P_i) \subset \mathcal{S}^N$. Agent $i$'s (partial) ordering $P_i \in \mathcal{S}$ over the set of alternative $2^K$ is then defined by "unanimity or consensus" among the rationales in $\mathcal{S}(P_i)$, in the sense that,

$$\forall X, Y \in 2^K, X \, P_i \, Y \iff X \, \tilde{P}_i \, Y \; \forall \tilde{P}_i \in \mathcal{S}(P_i). \tag{4.2}$$

In particular, when $\mathcal{S}(P_i)$ is a singleton for each agent $i \in N$, condition (4.2) reduces the analysis to the standard model with complete individual preferences over $2^K$ as in Barberà et al. (1991). In general, however, if there exists $\tilde{P}_i, \tilde{P}_i' \in \mathcal{S}(P_i)$ such that $X \, \tilde{P}_i \, Y$ and $Y \, \tilde{P}_i' \, X$ for some $X, Y \in 2^K$ and $i \in N$, then $X$ and $Y$ are incomparable to agent $i$ under $P_i$. The reason why this might happen in a social choice context includes cases where (i) each individual represents a collective of several agents with different rankings over the set of alternatives, or (ii) agents' complete preferences are not completely elicited due to limited communication. A case in point for the former would be a coalition of political parties ranking groups of legislators for a parliamentary committee. Although the coalition might agree upon the list of legislators suitable for the committee, party members might still have different views about the possible combinations of these candidates due to their party affiliation, which render control to the party over the committee (e.g., see Levy (2004)'s model of political parties). Similar problem is studied in computational social choice as accurately conveying an agent's complete preference relation over the set of alternative is proved to be very costly in terms of information and communication.

### 4.2.2 Axioms

In this section, we define several important strategic and non-strategic properties, or axioms, of rules. The first of these properties concerns with incentive compatibility. In domains with complete individual preferences, to guarantee truthful revelation, the planner usually demands from the voting rule an incentive compatibility property known as strategy-proofness. Roughly speaking, a voting rule is strategy-proof if, and only if, for each preference profile and every agent, truthfully reporting one's opinion leads to a social outcome that is weakly preferred to the outcome generated by any other untruthful declaration, regardless of what other agents report. In extending this property to the incomplete preference environment laid out above, one main challenge consists of determining how much consensus among the agents' rationales is demanded to justify an untruthful report. In this context, strategy-proofness admits at least two possible formulations, depending on whether an agent's lie is sustained by either one or all of its rationales. Borrowing the terminology from Heller (2012), we use the term "justifiable" for the former interpretation of strategy-proofness and define it as usual in negation to the concept of manipulability.

**Justifiable Manipulability**: There exists a preference profile $P \in \mathcal{S}^N$, $i \in N$ and $P'_i \in \mathcal{S}$ such that $f(P_i, P_{-i})\, \tilde{P}_i\, f(P'_i, P_{-i})$ *for some* $\tilde{P}_i \in \mathcal{S}(P_i)$.

If the rule $f$ is not justifiable manipulable, then it is *justifiable strategy-proof.*[3] Alternatively, we can write: the rule $f$ is justifiable strategy-proof if and only if for all $P \in \mathcal{S}^N$, $i \in N$ and $P'_i \in \mathcal{S}$,

$$f(P_i, P_{-i})\ P_i\ f(P'_i, P_{-i}) \text{ or } f(P_i, P_{-i}) = f(P'_i, P_{-i}).$$

The above definition is a very *cautious* property of incentive compatibility, that demands no profitable deviation from truth telling even if not all rationales compatible with one's opinion agree upon the desirability of a lie. Roughly speaking, the idea behind this axiom is to ensure that no agents has any justifications for lying about his true opinion. If the agents represent party coalitions, for

[3] Duggan and Schwartz (2000) uses similar concept termed *possible manipulation* based on the notion of first-order stochastic dominance to define nonmanipulability. Generally speaking, it requires that the agent find it unprofitable to deviate from truth telling for all possible lottery realizations over the set of assigned objects. Here, instead of using lottery realizations, we use the concept of justifications or rationales to determine the dominance of an alternative over another.

instance, then from the planner's viewpoint, justifiable strategy-proofness may be appropriate if he is not fully aware of how the agents resolve internally the conflict among the different rationales (party members' preferences), so that manipulation when a single party benefits from lying cannot be completely ruled out.

Obviously, in some settings it may be more appropriate to work with a more lax incentive requirement, that regards a voting rule as manipulable if a lie is justified or sustained not just by the profitability of the social outcome under one rationale, but under all of them. This demand for "(internal) consensus about the benefits of lying" brings about the second concept of manipulability and strategy-proofness.

**Manipulability**: There exists a preference profile $P \in \mathcal{S}^N$, $i \in N$ and $P_i' \in \mathcal{S}$ such that $f(P_i, P_{-i})\, \tilde{P}_i\, f(P_i', P_{-i})$ *for all* $\tilde{P}_i \in \mathcal{S}(P_i)$.

If the rule $f$ is not manipulable, then it is strategy-proof. Clearly, if the preference domain of each agent consists only of complete binary relations, then both above definitions converge to the concept of strategy-proofness over separable preferences employed in Barberà et al. (1991) and Ju (2003). In general, however, strategy-proofness is a weaker condition, in the sense that it is implied by the absent of justifiable manipulability. For the purpose of mechanism design, this is important as strategy-proofness offers the chance of exploiting the conflict among different selves or preferences of the agents to dismiss cases of manipulation that bring on disagreement and division instead of unity of purpose, enlarging the family of incentive compatible voting rules.

To see this in more details, we relate these two notions of strategy-proofness with the following two axioms, called monotonicity and independence, studied by Kasher and Rubinstein (1997) and Samet and Schmeidler (2003). First, monotonicity requires that when the set of desirable objects expand for all agents, then the set of socially chosen objects should also expand.

**Monotonicity**: For each pair of profiles $P$, $P' \in \mathcal{S}^N$ such that $D_i \subseteq D_i'$ for all $i \in N$, $f(P) \subseteq f(P')$.

The second axiom requires independence of decisions across objects. That is, the decision to include each object in the social outcome should depend only on agents' evaluations on the desirability of the object only.

**Independence**: For each object $x \in K$ and each pair of profiles $P$, $P' \in \mathcal{S}^N$

such that $N_x(P) = N_x(P')$, $x \in f(P)$ if and only if $x \in f(P')$.

## 4.3 Incentives

We first establish the relationship between the two notions of incentive compatibility with the axioms of monotonicity and independence.

**Proposition 4.1** *A voting rule is justifiable strategy-proof on the domain of incomplete and separable preferences if and only if it is monotonic and independent.*

**Proof.** (*Sufficiency*) Let $f$ be a monotonic and independent rule. Fix any $P \in \mathcal{S}^N$, $i \in N$, and $P'_i \in \mathcal{S}$. Consider the two following cases:

**Case 4.** If $D_i \cap D'_i \neq \emptyset$, then denote $D_i \cap D'_i = \hat{D}_i$ and $X = D_i \backslash D'_i$. Consider $\hat{P}_i \in \mathcal{S}$ such that $D(\hat{P}_i) = \hat{D}_i$. As $\hat{D}_i \subseteq D_i$, applying the monotonicity axiom to the pair of preference profiles $(P_i, P_{-i})$ and $(\hat{P}_i, P_{-i})$, we have $f(\hat{P}_i, P_{-i}) \subseteq f(P_i, P_{-i})$. On one hand, for all $x \in K$ such that $x \in D_i$ and $x \in \hat{D}_i$, independence axiom implies that $x \in f(\hat{P}_i, P_{-i}) \iff x \in f(P_i, P_{-i})$. On the other hand, for all $y \in K$ such that $y \notin D_i$ and $y \notin \hat{D}_i$, independence axiom implies that $y \notin f(\hat{P}_i, P_{-i}) \iff y \notin f(P_i, P_{-i})$. Thus, $f(P_i, P_{-i}) \setminus f(\hat{P}_i, P_{-i}) \subseteq X \subseteq D_i$. Then separability implies that

$$f(P_i, P_{-i}) \ P_i \ f(\hat{P}_i, P_{-i}), \text{ or } f(P_i, P_{-i}) = f(\hat{P}_i, P_{-i}). \tag{4.3}$$

Denote $Y = D'_i \setminus \hat{D}_i$. As $\hat{D}_i \subseteq D'_i$, applying monotonicity to the preference profiles $(P'_i, P_{-i})$ and $(\hat{P}_i, P_{-i})$, we have that $f(\hat{P}_i, P_{-i}) \subseteq f(P'_i, P_{-i})$. On one hand, for all $x \in K$ such that $x \in D'_i$ and $x \in \hat{D}_i$, independence property implies that $x \in f(\hat{P}_i, P_{-i}) \iff x \in f(P'_i, P_{-i})$. On the other hand, for all $y \in K$ such that $y \notin D'_i$ and $y \notin \hat{D}_i$, independence property implies that $y \notin f(\hat{P}_i, P_{-i}) \iff y \notin f(P'_i, P_{-i})$. Thus, $f(P'_i, P_{-i}) \setminus f(\hat{P}_i, P_{-i}) \subseteq Y$. Then, as $Y \cap D_i = \emptyset$, separability implies that

$$f(\hat{P}_i, P_{-i}) \ P_i \ f(P'_i, P_{-i}), \text{ or } f(\hat{P}_i, P_{-i}) = f(P'_i, P_{-i}). \tag{4.4}$$

Combining (4.3) and (4.4) and using the transitivity of the preference relation $P_i$, we have that either $f(P_i, P_{-i}) \ P_i \ f(P'_i, P_{-i})$, or $f(P_i, P_{-i}) = f(P'_i, P_{-i})$.

**Case 5.** If $D_i \cap D'_i = \emptyset$, then denote $D_i \cup D'_i = \check{D}_i$. Consider $\check{P}_i \in \mathcal{S}$ such that $D(\check{P}_i) = \check{D}_i$. Repeat the same argument in **Case 1** for the preference profiles $(P_i, P_{-i})$ and $(\check{P}_i, P_{-i})$, we have that either $f(P_i, P_{-i})\ P_i\ f(\check{P}_i, P_{-i})$ or $f(P_i, P_{-i}) = f(\check{P}_i, P_{-i})$. Similarly, repeat the argument for the profiles $(P'_i, P_{-i})$ and $(\check{P}_i, P_{-i})$, it follows that either $f(\check{P}_i, P_{-i})\ P_i\ f(P'_i, P_{-i})$ or $f(\check{P}_i, P_{-i}) = f(P'_i, P_{-i})$. Thus, by the transitivity property of the preference relation $P_i$, we have that either $f(P_i, P_{-i})\ P_i\ f(P'_i, P_{-i})$, or $f(P_i, P_{-i}) = f(P'_i, P_{-i})$.

To wrap up, since agent $i \in N$ and the preference profiles $(P_i, P_{-i}), (P'_i, P_{-i}) \in \mathcal{S}^N$ are arbitrarily chosen, **Case 1** and **Case 2** show that the voting rule $f$ (which by hypothesis satisfies monotonicity and independence) is justifiably strategy-proof.

(*Necessity*) We show first that justifiable strategy-proofness implies montonicity; and afterwards that strategy-proofness implies independence.

Appealing to the inductive arguments, considering any pair of preference profiles, $(P_i, P_{-i}), (P'_i, P_{-i}) \in \mathcal{S}^N$ such that $D'_i \subseteq D_i$, we need to show that $f(P'_i, P_{-i}) \subseteq f(P_i, P_{-i})$. Suppose, for contradiction, that there exists $x \in K$ such that $x \in f(P'_i, P_{-i})$ but $x \notin f(P_i, P_{-i})$. By definition of justifiable strategy-proofness, we have $f(P'_i, P_{-i})\ P'_i\ f(P_i, P_{-i})$ which implies $x \in D'_i$. By supposition, $x \in D'_i$ implies $x \in D_i$. Then we can construct a complete preference relation $\tilde{P}_i \in \mathcal{S}(P_i)$ such that for all $A, A' \subseteq K \setminus \{x\}$, $[A \cup \{x\}]\ \tilde{P}_i\ A'$. By construction, $f(P_i, P_{-i})\ \tilde{P}_i\ f(P'_i, P_{-i})$, which violates the justifiable strategy-proofness property. Repeat the induction step for all $i \in N$, we have justifiable strategy-proofness implies monotonicity.

Again, using the inductive argument, consider any object $x \in K$ and any pair of preference profiles $(P_i, P_{-i}), (P'_i, P_{-i}) \in \mathcal{S}^N$ such that $x \in D_i \iff x \in D'_i$, we need to show that $x \in f(P_i, P_{-i}) \iff x \in f(P'_i, P_{-i})$. Suppose, for contradiction, that $x \in f(P'_i, P_{-i})$ but $x \notin f(P_i, P_{-i})$. By the property of justifiable strategy-proofness, we have $f(P'_i, P_{-i})\ P'_i\ f(P_i, P_{-i})$ which implies $x \in D'_i$. By supposition, $x \in D'_i$ if and only if $x \in D_i$. Then we can construct a complete preference relation $\tilde{P}_i \in \mathcal{S}(P_i)$ such that for all $A, A' \subseteq K \setminus \{x\}$, $[A \cup \{x\}]\ \tilde{P}_i\ A'$. By construction, $f(P_i, P_{-i})\ \tilde{P}_i\ f(P'_i, P_{-i})$, which violates the justifiable strategy-proofness property. Repeat the induction step for all $i \in N$, we have justifiable strategy-proofness implies independence. ∎

The characterization given in Proposition 4.1 is tight, in the sense that a rule satisfying either monotonicity or independence alone does not necessarily satisfy justifiable strategy-proofness.[4] This relationship appears quite often in the strategy-proof analysis with complete preferences, as is the case of Nehring and Puppe (2007) for generalized single-peaked preferences, and of Ju (2003) for separable domain with nonlinear preferences. Similar to these cases, in our framework with incomplete preferences, Proposition 4.1 plays a pivotal role in unveiling the family of incentive compatible voting rules.

The next proposition establishes the relationship between justifiable strategy-proofness and standard notion of nonmanipulability.

**Proposition 4.2** *A voting rule is justifiable strategy-proof on the domain of incomplete and separable preferences if and only if it satisfies strategy-proofness and independence.*

**Proof.** (*Sufficiency*) By Proposition 4.1, justifiable strategy-proofness is equivalent to the axioms of monotonicity and independence. Thus, to show that strategy-proofness and independence implies justifiable strategy-proofness, it is sufficent to show that strategy-proofness and independence implies monotonicity.

Consider any pair of preference profiles, $(P_i, P_{-i}), (P'_i, P_{-i}) \in \mathcal{S}^N$ such that $D_i = D'_i \cup \{x\}$ for some $x \in K \setminus D'_i$, we need to show that $f(P'_i, P_{-i}) \subseteq f(P_i, P_{-i})$. As $N_y(P'_i, P_{-i}) = N_y(P_i, P_{-i})$ for all $y \in K \setminus \{x\}$, by independence axiom, $y \in f(P'_i, P_{-i})$ if and only if $y \in f(P_i, P_{-i})$. Suppose, for contradiction, that $x \in f(P'_i, P_{-i})$ but $x \notin f(P_i, P_{-i})$. Then $f(P'_i, P_{-i}) = f(P_i, P_{-i}) \cup \{x\}$. As $x \in D_i$, we have $f(P'_i, P_{-i})\ \tilde{P}_i\ f(P_i, P_{-i})$ for all $\tilde{P}_i \in \mathcal{S}(P_i)$, contradiction. Thus, $f(P'_i, P_{-i}) \subseteq f(P_i, P_{-i})$.

Repeate the induction step for all $i \in N$, we have for every pair of preference profiles $(P_i)_{i\in N}, (P'_i)_{i\in N} \in \mathcal{S}^N$, if $D_i \subseteq D'_i$ for all $i \in N$, then $f(P_i, P_{-i}) \subseteq f(P'_i, P'_{-i})$.

(*Necessity*) By Proposition 4.1, justifiable strategy-proofness implies independence. By definition, manipulability implies justifiable manipulability. Thus, justifiable strategy-proofness implies strategy-proofness and independence. ∎

[4]For example, fix an agent $i \in N$ and consider a rule $f$ such that for all $(P_i, P_{-i}) \in \mathcal{S}^N$ and each $x \in K$, $x \in f(P_i, P_{-i}) \iff x \notin D_i$. This rule is independent, but it violates monotonicity and justifiable strategy-proofness. Alternatively, redefine $f$ in such a way that for some $i \in N$ and $x \in K$, $f(P_i, P_{-i}) = K$ if $x \in D_i$; and $f(P_i, P_{-i}) = \emptyset$ otherwise. Clearly, $f$ is monotonic but does not satisfy independence. Moreover, for any rationale $\tilde{P}_i \in \mathcal{S}(P_i)$ such that $D_i = \{x\}$ and $\emptyset\ \tilde{P}_i\ K$, $f$ is justifiable manipulable.

Proposition 4.2 provides an insight on the environment where justifiable strategy-proofness can replace the standard notion of incentive compatibility. In the context where there is little correlation among the consumption of each object, the decision to include an object is not compromised by the social decision on another and independence axioms is trivially satisfied. This assumption usually appears in the literature of group identification or committee selection. In such cases, the two notions of strategy-proofness converge and the characterization of strategy-proof voting rule coincides. On the other hand, when there is an exogenous and common knowledge relationship among objects, the set of strategy-proof voting rules expands and it is no longer characterized by the axioms of monotonicity and independence. The arguments for such cases include objects with positive/negative externalities, complementary and substitute goods or restricted domain of social alternatives. Here is an example where a strategy-proof voting rule can be justifiable manipulable.

**Example** Let $f^U$ be the unanimity rule that chooses an object if and only if all agents consider it desirable. Let $f^L$ be a rule that choose a single object $x \in \cup_{i \in N} D_i$ according to a predetermined lexicographical order of $K$. Consider the following voting rule $f^*$:

$$f^* = \begin{cases} f^U & \text{if there exists } x \in K \text{ such that } N_x(P) = N; \\ f^L & \text{if } N_x(P) \neq N \text{ for all } x \in K. \end{cases} \tag{4.5}$$

By Proposition 4.1, $f^*$ does not satisfy either monotonicity or independence axioms, thus, it is justifiable manipulable. However, it is trivial that the rule satisfies strategy-proofness.

To characterize the set of justifiable strategy-proof rules, first define the following notation and terminology. Define a committee for $x \in K$ as a pair $\mathcal{C}_x = (N, \mathcal{W}_x)$, where $\mathcal{W}_x$ is a set of coalitions of $N$ which satisfies the condition that for each pair $M$, $M' \subseteq N$, if $M \in \mathcal{W}_x$ and $M \subseteq M'$, then $M' \in \mathcal{W}_x$. Let $\mathcal{C} = (\mathcal{C}_x)_{x \in K}$ be a committee profile. The voting-by-committees rule with respect to $\mathcal{C}$ is defined as follows: for each profile $P = (P_i)_{i \in N} \in \mathcal{S}^N$ and every object $x \in K$, $x \in f(P)$ if, and only if, $N_x(P) \in \mathcal{W}_x$. In words, a voting-by-committees rule is one that chooses an object $x$ if, and only if, the set of agents who consider $x$ a desirable object constitutes a (winning) coalition within $x$'s decision committee $\mathcal{W}_x$. Denote $\mathcal{F}^C$ the set of voting-by-committees rules.

**Theorem 4.1** *A voting rule is justifiably strategy-proof on the domain of incomplete and separable preferences if and only if it is a voting-by-committees rule.*

The proof of Theorem 4.1 is simiar to that of Barberà et al. (1991) for strategy-proof voting schemes that satisfy "tops only" property.

**Proof.** *(Necessity)* Each voting-by-committees rule is monotonic and independent. Thus, by Proposition 4.1, the rule is justifiably strategy-proof.

*(Sufficiency)* Let $f$ be a justifiably strategy-proof rule. By definition, $f$ is a voting-by-committees rule is and only if for any $x \in K$ and any preference profiles $P \in \mathcal{S}^N$, there exists a profile of committees $\mathcal{C}_x = (N, \mathcal{W}_x)$ such that

$$x \in f(P) \iff N_x(P) \in \mathcal{W}_x. \tag{4.6}$$

We will show that for each justifiably strategy-proof voting rule, we can define a generalized committee that specify all the winning coalition for each object $x \in K$. Fix any object $x \in K$. Let $\mathcal{S}(f, x) \subseteq \mathcal{S}^N$ be a set of preference profiles such that for all $P \in \mathcal{S}(f, x)$, $x \in f(P)$. We defined object $x$'s generalized committee $\mathcal{W}_x$ for each of the following cases:

If $\mathcal{S}(f, x) = \emptyset$. That is, for all $P \in \mathcal{S}^N$, $x \notin f(P)$. Set $\mathcal{W}_i = \emptyset$.

If $\mathcal{S}(f, x) \neq \emptyset$. That is, there exists $P \in \mathcal{S}^N$ such that $x \in f(P)$.

Consider $\mathcal{W}_x = \{N_x(P) \subseteq N : P \in \mathcal{S}(f, x)\}$. Let $M$ be an element of $\mathcal{W}_x$ and $M' \supseteq M$. By definition, $\mathcal{W}_x$ is a committee if and only if $M' \in \mathcal{W}_x$. That is, we need to show that there exists a preference profile $P' \in \mathcal{S}^N$ such that $M' = N_x(P')$ and $x \in f(P')$. By supposition, $M = N_x(P)$ for some $P \in \mathcal{S}^N$ then we can construct $P' \in \mathcal{S}^N$ as follows, $D'_i = K$ for all $i \in M' \setminus M$ and $D'_j = D_j$ otherwise. Then by monotonicity, as $D_i \subseteq D'_i$ for all $i \in N$, $f(P) \subseteq f(P')$. Since $x \in f(P)$, then $x \in f(P')$ and $M' \in \mathcal{W}_x$. Thus, $\mathcal{C}_x = (N, \mathcal{W}_x)$ constitutes a committee for each $x \in K$.

Finally, we will show that for any $x \in K$ and $P \in \mathcal{S}^N$, $x \in f(P) \iff N_x(P) \in \mathcal{W}_x$. The "only if" part is trivially satisfied by construction of $\mathcal{W}_i$. On the other hand, if $N_x(P) \in \mathcal{W}_x$, then there exists $P' \in \mathcal{S}^N$ such that $N_x(P) = N_x(P')$ and $x \in f(P')$. As $x \in D_i \iff x \in D'_i$ for all $i \in N$, apply independence axiom, we have $x \in f(P')$ implies $x \in f(P)$.

Apply the same argument for all $x \in K$, we obtain the profile of committees $(\mathcal{C}_x)_{x \in K}$ satisfying the condition. Therefore, $f$ is justifiably strategy-proof if and only if it is voting-by-committee rules. ∎

Barberà et al. (1991) provides a similar characterization of voting-by-committees rules but over the domain of complete and separable preferences. More recently, Hatsumi et al. (2014) shows that on the set of complete individual preferences, separability constitutes a maximal domain condition for the existence of nontrivial rules satisfying strategy-proofness and no-vetoer. Theorem 4.1 contributes to this literature showing that the maximal domain can in fact be enlarged by including incomplete preference relations that satisfy separability and for which the set of best objects is well-defined. To be more precise, it shows that over these preference domains voting-by-committees rules remains the only family of incentive compatible voting rules.

## 4.4 Efficiency

To identify the set of optimal voting-by-committees rules, we first use the notion of efficiency which is defined using ordinal information and study its implications. As voting-by-committees rules guarantee to satisfy the incentive compatible constraint, we can safely evaluate the voting rules in this family based on their ex post efficiency. Consider a preference profile $P \in \mathcal{S}^N$ and rationale profile $\tilde{P} \in \mathcal{S}(P)$. Recall that for each pair alternatives $X$, $Y \in 2^K$, $X$ *Pareto dominates* $Y$ for $\tilde{P}$ if for each $i \in N$, $X \, \tilde{P}_i \, Y$. We say that an alternative $X$ is *Pareto efficient* for $\tilde{P}$ if and only if it is not Pareto dominated for $P$ by any other alternatives. In our context of incomplete and separable preference domain, the notion of Pareto efficiency means that, if there is no alternative $Y$ that Pareto dominates $X$, then there exist a least one justification profile $\tilde{P} \in \mathcal{S}(P)$ such that $X$ is Pareto efficient for $\tilde{P}$. A voting rule is efficient if for all preference profile $P \in \mathcal{S}^N$, the outcome chosen under such rule is Pareto efficient.

**Efficiency**: For all preference profile $P \in \mathcal{S}^N$, there does not exist $X \in 2^K$ such that $X \, P_i \, f(P)$ for all $i \in N$.[5]

Notice that as the preference relation $P_i$ contains many possible justifications $\tilde{P}_i$, our notion of Pareto dominance is strong and hence, the efficiency restriction is weak. In fact, the only restriction that this notion of efficiency imposes on the structure of the committees is that of unanimity: if all agents agree that an

[5]This notion of efficiency is similar to "ordinal efficiency" which is defined on (partial) preferences over lotteries that are obtained from preferences over sure outcomes by applying first-order stochastic dominance (see Danan et al. 2016, Carroll 2010).

object is desirable, then such object should be included in the social outcome.

**Proposition 4.3** *A voting-by-committees rule $f \in \mathcal{F}^C$ is efficient on the domain of incomplete and separable preferences if and only if $\mathcal{W}_x \neq \{\emptyset, \{\emptyset\}\}$ for all $x \in K$.*

Proposition 4.3 is encouraging in the sense that, different from Barberà et al. (1991), in our domain of incomplete preferences, the set of efficient voting-by-committees rules expands and contains more than just dictatorial rules. However, it also means that the notion of efficiency is too weak to narrow down $\mathcal{F}^C$ in a significant way.

## 4.5 Cardinal Criterion

Once the set of efficient outcomes is identified, one may wish to select among them the "optimal outcome" using different cardinal criterion. It requires first, a utility representation that preserves all the information of an incomplete and separable preference and second, a method to aggregate those utility functions to evaluate the performance of different voting-by-committees rules. There are two main notions of utility representation for an incomplete preference relation $P$ on the set of alternatives $2^K$. The first approach is to use Richter-Peleg utility representation which is a real function over $2^K$ constructed using available information of the relation $P$. However, one of the main limitations of Richter-Peleg representation is that it does not characterize the original preference relation but rather extends it to a complete pre-order that is representable in the usual sense. Alternatively, Ok (2002) extends upon the standard notion of Debreu's utility representation and show that an incomplete and transitive binary relation $P_i \in \mathcal{S}$ over $2^K$ admits a *multi-utility representation* if and only if there exists a convex set $\mathcal{U}_i$ of real mapping on $2^K$ such that for any social alternatives $X, Y \in 2^K$,

$$X \ P_i \ Y \iff u_i(X) > u_i(Y) \text{ for all } u_i \in \mathcal{U}_i. \tag{4.7}$$

The first advantage of multi-utility representation is that it preserves all the information on the situations in which the agent is actually indecisive. Instead of a single real function, an incomplete preference relation is represented by a vector-valued utility function the range of which is contained in some partially ordered linear space. This multi-utility representation is also consistent with the justifiable choice approach of Heller (2012) in the sense that each complete and separable rationale can be represented by a real utility function in $\mathcal{U}_i$.

More specifically, consider an agent $i \in N$ with preference relation $P_i \in \mathcal{S}$. Applying the concept of multi-utility representation to our environment, then any incomplete and separable preference relation $P_i \in \mathcal{S}$ can be represented by the following utility representation: for any subset $X \subseteq K$, the utility of an agent endowed with $\tilde{P}_i \in \mathcal{S}(P_i)$ get from the social outcome $X$ is given by,

$$u_i(P_i, X) = -\sum_{x \in K} [\alpha_i^x \cdot \mathbf{I}_{X \setminus D_i}(x) + \beta_i^x \cdot \mathbf{I}_{D_i \setminus X}(x)]. \tag{4.8}$$

with $\alpha_i^x, \beta_i^x \in \mathbb{R}_{++}$, where the first indicator function is to determine if $x$ is an undesirable object included in the social outcome $A$, that is,

$$\mathbf{I}_{X \setminus D_i}(x) = \begin{cases} 1 & \text{if } x \in X \setminus D_i; \\ 0 & \text{otherwise,} \end{cases}$$

and the second indicator function is to determine if $x$ belongs to the set of agent $i$'s desirable objects excluded from the social outcome $A$, that is,

$$\mathbf{I}_{D_i \setminus X}(x) = \begin{cases} 1 & \text{if } x \in D_i \setminus X; \\ 0 & \text{otherwise.} \end{cases}$$

If we consider each subset $X$ of $K$ as a point in multidimensional space $X \in \{0,1\}^K$, then our utility representation basically measure the weighted Euclidean distance between the social decision to each individual's most preferred alternative. Let $\boldsymbol{\alpha}_i = (\alpha_i^1, \ldots, \alpha_i^k)$, $\boldsymbol{\beta}_i = (\beta_i^1, \ldots, \beta_i^k)$ and define $H(P_i) \subseteq \mathbb{R}_{++}^{2k}$ agent $i$'s (convex) set of admissible parameter vectors such that each $\tilde{P}_i \in \mathcal{S}(P_i)$ can be represented by $u_i$ under a set of parameters $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i) \in H(P_i)$. Formally, for any alternatives $X, X' \subseteq K$,

$$\begin{aligned} H(P_i) \equiv \{(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i) \in \mathbb{R}_{++}^{2k} : \text{ for all } \tilde{P}_i \in \mathcal{S}(P_i) \text{ there exists } (\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i) \in H(P_i) \\ \text{such that } u_i(P_i, X) > u_i(P_i, X') \iff X \; \tilde{P}_i \; X'\}. \end{aligned} \tag{4.9}$$

Apply this utility function for all preference relation $\tilde{P}_i \in \mathcal{S}(P_i)$, we can derive the multi utility representation for the preference relation $P_i$ as follows: for any alternatives $X, X' \subseteq K$,

$$X\ P_i\ X' \iff u_i(P_i, X) > u_i(P_i, X') \quad \forall(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i) \in H(P_i). \tag{4.10}$$

That is, for any agent $i \in N$, the convex set $\mathcal{U}_i$ can be defined as:

$$\mathcal{U}_i \equiv \{u_i(P_i) \,:\, (\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i) \in H(P_i)\}. \tag{4.11}$$

For simplicity, we assume that $\alpha_i^x, \beta_i^x \in [\underline{\varepsilon}_i^x, \bar{\varepsilon}_i^x] \subset \mathbb{R}_{++}$, then the (convex) set of admissible parameters $H(P_i) = \prod_{x\in K}[\underline{\varepsilon}_i^x, \bar{\varepsilon}_i^x] \times \prod_{x\in K}[\underline{\varepsilon}_i^x, \bar{\varepsilon}_i^x]$. Denote $u = (u_i)_{i\in N}$ a utility profile in $\prod_{i\in N}\mathcal{U}_i$ and $H(P) = \prod_{i\in N} H(P_i)$ the set of parameter profile of the society.

Having defined the multi-utility representation for each individual's preference relation, now we need to aggregate individual utilities in a way that is consistent with Paretian principle. Danan, Gajdos and Tallon (2015) show that Harsanyi's aggregation theorem remains possible, even when individuals admit incomplete preferences over alternatives. Specifically, it says that a social preference satisfies Pareto preference if and only if it admits a multi-utility representation that is a set of utilitarian aggregations of individual utility functions. In this paper, we define the (utilitarian) social welfare function $\mathcal{U}_0$ as the sum of each individual's multi-utility functions, that is, for any preference profile $P \in \mathcal{S}^N$ and social alternative $X \subseteq K$, we write,

$$\mathcal{U}_0(P, X) = \Big\{ \sum_{i\in N} u_i(P_i, X) \,:\, u \in \prod_{i\in N} \mathcal{U}_i \Big\}, \tag{4.12}$$

with each social welfare function $u_0 = \sum_{i\in N} u_i(P_i, X) \in \mathcal{U}_0$ corresponds to a profile of parameters $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)_{i\in N} \in H(P)$.

While the multi-utility representation allows us to map ordinal information into a real-valued space and take aggregation in a manner that is consistent with Pareto principle, it does not provide any further implications on how voting rules can be evaluated. Virtually any consistent utility profile $u \in \prod_{i\in N}\mathcal{U}_i$ can be interpreted as evaluating the different voting-by-committees rules with varying degrees of justice by the social planner. The problem of the social planner is similar to that of an agent who has to make decisions under the uncertainty about the state of the world. Assuming that we have information about $H(P_i)$ but not the value or probabilistic distribution of $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$, one can define the optimality the following sense:

**Optimality**: For any preference profile $P \in \mathcal{S}$ endowed with multi-utility representations $\prod_{i \in N} \mathcal{U}_i$ and any $f' \in \mathcal{F}^C$,

$$V(P, f) \geq V(P, f')$$

where $V : \mathcal{S}^N \times \mathcal{F}^C \to \mathbb{R}$ is the social planner's valuation of different voting rules.

Before characterizing the set of optimal voting-by-committees for different valuation functions $V$, we first define a class of rules in $\mathcal{F}^C$ that will play an important role in the analysis below. A voting-by-committees rule $f \in \mathcal{F}^C$ is weighted majority rule if for any object $x \in K$, there are strictly positive numbers $(w_1^x, \ldots, w_n^x)$ such that

$$\mathcal{W}_x = \{M \subseteq N : \sum_{i \in M} w_i^x \geq \sum_{i \notin M} w_i^x\}. \tag{4.13}$$

If $w_1^x = \cdots = x_n^x$ then it is a simple majority rule which selects an object if and only if it is considered desirable for the majority of individuals. Note that the inequality in (4.13) is lax for the purpose of assigning a tie-breaking rule: from an ordinal viewpoint, as any non-degenerate voting-by-committees rule is (justifiable) strategy-proof and efficient, the tie-breaking only concerns the cardinal optimality and can be determined arbitrarily, depending on the nature of the objects considered.

### 4.5.1 Optimism-pessimism

One of the most well-known criterion for decision-making under uncertainty is the Hurwicz's criterion, which selects an optimal act using the weighted average of minimum and maximum payoffs. Specifically, in our context, the valuation function $V_H$ takes the form:

$$V^H(P, f) = \lambda \min_{\prod_{i \in N} \mathcal{U}_i} \sum_{i \in N} u_i(P_i, f(P)) + (1 - \lambda) \max_{\prod_{i \in N} \mathcal{U}_i} \sum_{i \in N} u_i(P_i, f(P)), \tag{4.14}$$

where the coefficient $\lambda \in [0, 1]$ can be interpreted as the degree of caution with which the voting-by-committees rule $f$ is evaluated. Given the convexity of $\mathcal{U}_0$, for any $u_0 \in \mathcal{U}_0$, there exist a unique $\lambda$ such that $u_0(P, f(P)) = V^H(P, f)$. The most cautious rule corresponds to the case $\lambda = 1$ where each voting rule is evaluated under the maximin principle. On the hand, when $\lambda = 0$, each voting rule is evaluated based on their maximum social welfare, which corresponds to the maximax principle.

**Theorem 4.2** *A voting-by-committees rule* $f^H : \mathcal{S}^N \to 2^K$ *is optimal under Hurwicz criterion if it is weighted majority rule with* $w_i^x = \lambda\bar{\varepsilon}_i^x + (1-\lambda)\underline{\varepsilon}_i^x$ *for all* $i \in N$.

**Proof.** Fix any preference profile $P \in \mathcal{S}^N$. First, consider any $x \in f^H(P)$ and $f' \in \mathcal{F}^C$ such that $f'(P) = f^H(P) \setminus \{x\}$. As the voting-by-committees rule $f^H$ is optimal under Hurwicz criterion, we have

$$V^H(P, f^H) - V^H(P, f) \geq 0 \tag{4.15}$$

$$\iff \big[\lambda \min_{H(P)} \sum_{i \notin N_x(P)} (-\alpha_i^x) + (1-\lambda) \max_{H(P)} \sum_{i \notin N_x(P)} (-\alpha_i^x)\big]$$
$$-\big[\lambda \min_{H(P)} \sum_{i \in N_x(P)} (-\beta_i^x) + (1-\lambda) \max_{H(P)} \sum_{i \in N_x(P)} (-\beta_i^x)\big] \geq 0 \tag{4.16}$$

$$\iff \sum_{i \in N_x(P)} \big[\lambda\bar{\varepsilon}_i^x + (1-\lambda)\underline{\varepsilon}_i^x\big] - \sum_{i \notin N_x(P)} \big[\lambda\bar{\varepsilon}_i^x + (1-\lambda)\underline{\varepsilon}_i^x\big] \geq 0 \tag{4.17}$$

Similar arguments can be applied for any $x \notin f^H(P)$ by considering $f' \in \mathcal{F}^C$ such that $f'(P) = f^H(P) \cup \{x\}$. Thus, a voting-by-committees rule is optimal under Hurwicz's $\lambda$-criterion if it is a weighted quota rule with $w_i^x = \lambda\bar{\varepsilon}_i^x + (1-\lambda)\underline{\varepsilon}_i^x$ for all $i \in N$. ∎

The intuition behind the committees defined in Theorem 4.2 is that under Hurwicz's criterion, the social planner will apply the same level of caution when evaluating the social welfare to each individuals' preferences. Thus, each individual preference will be evaluated under the same precautionary principle (same $\lambda$) and the optimal rule is simply a weighted majority rule with caution-adjusted weight assigned to each individual.

### 4.5.2 Minimax regret

Consider the following valuation function:

$$V^R(P, f) = \max_{\prod_{i \in N} \mathcal{U}_i} \big[\sum_{i \in N} u_i(P_i, f(P)) - \max_{X \subseteq K} \sum_{i \in N} u_i(P, X)\big]. \tag{4.18}$$

This valuation function reflects the maximum utility loss experienced by selecting the social alternative $f(P)$ instead of choosing a welfare-maximizing outcome $X \subseteq K$ over all possible rationale profile $(u_i)_{i\in N} \in \prod_{i\in N} \mathcal{U}_i$. It can also be interpreted as the maximum regret of a social planner by implementing voting-by-committees rule $f \in \mathcal{F}^C$. Then the optimal voting-by-committees $f^R$ induced by valuation function $V^R$ chooses a feasible outcome that minimize the maximum regret of the social planner with respect to all possible realizations of utility profile.

**Theorem 4.3** *A voting-by-committees rule $f^R : \mathcal{S}^N \to 2^K$ is optimal under minimax regret criterion if it is weighted majority rule with*

$$w_i^x = \bar{\varepsilon}_i^x + \underline{\varepsilon}_i^x \quad \textit{for all}\, i \in N;$$

$$\textit{or} \qquad w_i^x = \begin{cases} \underline{\varepsilon}_i^x & \textit{for}\, i \in N_x(P), \\ \bar{\varepsilon}_i^x & \textit{otherwise.} \end{cases}$$

That is, the committees for any object $x \in K$ can be defined as follows:

$$\mathcal{W}_x^R = \left\{ M \subseteq N : \sum_{i\in M} (\underline{\varepsilon}_i^x + \bar{\varepsilon}_i^x) \geq \sum_{i\notin M} (\underline{\varepsilon}_i^x + \bar{\varepsilon}_i^x) \right\} \cup \left\{ M \subseteq N : \sum_{i\in M} \underline{\varepsilon}_i^x \geq \sum_{i\notin M} \bar{\varepsilon}_i^x \right\}. \tag{4.19}$$

**Proof.** Fix any preference profile $P \in \mathcal{S}^N$. It is trivial to check that

$$x \in \arg\max_{X\subseteq K} \sum_{i\in N} u_i(P_i, X) \iff \sum_{i\in N_x(P)} \beta_i^x \geq \sum_{i\notin N_x(P)} \alpha_i^x. \tag{4.20}$$

First, consider any $x \in f^R(P)$ and $f' \in \mathcal{F}^C$ such that $f'(P) = f^R(P) \setminus \{x\}$. As the voting-by-committees rule $f^R$ is optimal under minimax regret criterion, we have: for all $f \in \mathcal{F}^C$

$$V^R(P, f^R) - V^R(P, f) \geq 0. \tag{4.21}$$

Consider the three following cases:

**Case 1.** $\sum_{i\in N_x(P)} \underline{\varepsilon}_i^x > \sum_{i\notin N_x(P)} \bar{\varepsilon}_i^x$. Then $x \in \arg\max_{X\subseteq K} \sum_{i\in N} u_i(P_i, X)$ for all utility profile $(u_i)_{i\in N} \in \prod_{i\in N} \mathcal{U}_i$. It is trivially followed that $V^R(P, f^R) - V^R(P, f') > 0$.

**Case 2.** $\sum_{i\in N_x(P)} \bar{\varepsilon}_i^x < \sum_{i\notin N_x(P)} \underline{\varepsilon}_i^x$. Then $x \notin \arg\max_{X\subseteq K} \sum_{i\in N} u_i(P_i, X)$ for all utility profile $(u_i)_{i\in N} \in \prod_{i\in N} \mathcal{U}_i$. It is trivially followed that $V^R(P, f^R) -$

$V^R(P, f') < 0$.

**Case 3.** $\sum_{i \notin N_x(P)} \underline{\varepsilon}_i^x \leq \sum_{i \in N_x(P)} \beta_i^x \leq \sum_{i \notin N_x(P)} \bar{\varepsilon}_i^x$ for some $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)_{i \in N} \in H(P)$. Then we can rewrite (4.21) as follows:

$$\begin{aligned} & V^R(P, f^R) - V^R(P, f') \geq 0 \\ \iff & \Big[ \sum_{i \in N_x(P)} \underline{\varepsilon}_i^x - \sum_{i \notin N_x(P)} \bar{\varepsilon}_i^x \Big] - \Big[ \sum_{i \notin N_x(P)} \underline{\varepsilon}_i^x - \sum_{i \in N_x(P)} \bar{\varepsilon}_i^x \Big] \geq 0 \\ \iff & \sum_{i \in N_x(P)} \big[ \underline{\varepsilon}_i^x + \bar{\varepsilon}_i^x \big] - \sum_{i \notin N_x(P)} \big[ \underline{\varepsilon}_i^x + \bar{\varepsilon}_i^x \big] \geq 0 \end{aligned}$$

Similar arguments can be applied for any $x \notin f^R(P)$ by considering $f' \in \mathcal{F}^C$ such that $f'(P) = f^R(P) \cup \{x\}$. Thus, $x$ is included in the minimax regret social outcome $f^R$ if and only if,

$$x \in f^R(P) \iff \begin{cases} \sum_{i \in N_x(P)} \underline{\varepsilon}_i^x & \geq \sum_{i \notin N_x(P)} \bar{\varepsilon}_i^x; \quad \text{or} \\ \sum_{i \in N_x(P)} \big[ \underline{\varepsilon}_i^x + \bar{\varepsilon}_i^x \big] & \geq \sum_{i \notin N_x(P)} \big[ \underline{\varepsilon}_i^x + \bar{\varepsilon}_i^x \big]. \end{cases} \tag{4.22}$$

Define a committee profile $\mathcal{W}_x^R$ for each $x \in K$ according to equation (4.22). Then by construction, $f^R \in \arg\max_{f \in \mathcal{F}^C} V^R(P, f)$ for all $P \in \mathcal{S}^N$. ∎

The intuition behind the committees defined in Theorem 4.3 is that a regret averse social planner will take into account the incompleteness in individual's preference domain and evaluate the society's marginal utility at the extreme values ($\underline{\varepsilon}_i^x$ and $\bar{\varepsilon}_i^x$). The first committee structure requires that if the maximum regret generated by having $x$ included in the social outcome is smaller than the maximum regret from excluding $x$, then $x$ should be included. The second committee structure requires that as maximum regret is greater or equal than 0, if the maximum regret of including $x$ is less than or equal to 0, then having $x$ excluded will be weakly worse than including it.

### 4.5.3 Distortion

Caragiannis and Procaccia (2011) propose the notion of distortion to evaluate the performance of voting rules in an environment where there is limited communication between voters and the social planner. The distortion of a voting rule is the worst-case ratio of the social welfare of the alternative selected by $f$ to that of the utilitarian optimal alternative. In our context, the notion of distortion corresponds to the following valuation function:

$$V^D(P,f) = \max_{\prod_{i\in N}\mathcal{U}_i} \log\left[\frac{\max_{X\subseteq K}\sum_{i\in N} u_i(P_i,X)}{\sum_{i\in N} u_i(P_i,f(P))}\right] \tag{4.23}$$

The distortion minimizing voting-by-committees $f^D \in \mathcal{F}^C$ must choose a feasible outcome that minimize the distortion of the voting rule with respect to all possible realizations of utility profile.

**Theorem 4.4** *A voting-by-committees rule $f^D : \mathcal{S}^N \to 2^K$ minimize the distortion if it is weighted majority rule with*

$$w_i^x = \begin{cases} \bar{\varepsilon}_i^x \sum_{i\in N_x(P)} \underline{\varepsilon}_i^x & \text{for } i \in N_x(P), \\ \underline{\varepsilon}_i^x \sum_{i\notin N_x(P)} \bar{\varepsilon}_i^x & \text{otherwise.} \end{cases};$$

$$\text{or} \qquad w_i^x = \begin{cases} \underline{\varepsilon}_i^x & \text{for } i \in N_x(P), \\ \bar{\varepsilon}_i^x & \text{otherwise.} \end{cases}$$

That is, the committees for any object $x \in K$ can be defined as follows:

$$\mathcal{W}_x^D = \left\{M \subseteq N : \sum_{i\in M}\underline{\varepsilon}_i^x \sum_{i\in M}\bar{\varepsilon}_i^x \geq \sum_{i\notin M}\underline{\varepsilon}_i^x \sum_{i\notin M}\bar{\varepsilon}_i^x\right\} \cup \left\{M \subseteq N : \sum_{i\in M}\underline{\varepsilon}_i^x \geq \sum_{i\notin M}\bar{\varepsilon}_i^x\right\}. \tag{4.24}$$

**Proof.** Fix any preference profile $P \in \mathcal{S}^N$. By the independence of voting-by-committees, we can evaluate the decision to include an object in the social outcome independently. First, consider any $f^D(P) = \{x\}$ and $f' \in \mathcal{F}^C$ such that $f'(P) = \emptyset$. The proof is similar to the case with minimax regret criteria. The voting-by-committees rule $f_D$ minimizes the distortion if and only if for all $f \in \mathcal{F}^C$,

$$V^D(P,f^D) - V^D(P,f) \geq 0. \tag{4.25}$$

Consider the three following cases:

**Case 1.** $\sum_{i\in N_x(P)}\underline{\varepsilon}_i^x > \sum_{i\notin N_x(P)}\bar{\varepsilon}_i^x$. Then $x \in \arg\max_{X\subseteq K}\sum_{i\in N} u_i(P_i,X)$ for all utility profile $(u_i)_{i\in N} \in \prod_{i\in N}\mathcal{U}_i$. It is trivially followed that $V^D(P,f^D) - V^D(P,f') > 0$.

**Case 2.** $\sum_{i\in N_x(P)}\bar{\varepsilon}_i^x < \sum_{i\notin N_x(P)}\underline{\varepsilon}_i^x$. Then $x \notin \arg\max_{X\subseteq K}\sum_{i\in N} u_i(P_i,X)$ for all utility profile $(u_i)_{i\in N} \in \prod_{i\in N}\mathcal{U}_i$. It is trivially followed that $V^D(P,f^D) - V^D(P,f') < 0$.

**Case 3.** $\sum_{i\notin N_x(P)} \underline{\varepsilon}_i^x \leq \sum_{i\in N_x(P)} \beta_i^x \leq \sum_{i\notin N_x(P)} \bar{\varepsilon}_i^x$ for some $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)_{i\in N} \in H(P)$. Then we can rewrite (4.25) as follows:

$$\begin{aligned} & V^D(P, f^D) - V^D(P, f') \geq 0 \\ \iff \quad & \frac{\sum_{i\in N_x(P)} \underline{\varepsilon}_i^x}{\sum_{i\notin N_x(P)} \bar{\varepsilon}_i^x} - \frac{\sum_{i\notin N_x(P)} \underline{\varepsilon}_i^x}{\sum_{i\in N_x(P)} \bar{\varepsilon}_i^x} \geq 0 \\ \iff \quad & \sum_{i\in N_x(P)} \underline{\varepsilon}_i^x \sum_{i\in N_x(P)} \bar{\varepsilon}_i^x - \sum_{i\notin N_x(P)} \underline{\varepsilon}_i^x \sum_{i\notin N_x(P)} \bar{\varepsilon}_i^x \geq 0 \end{aligned}$$

Thus, a voting-by-committees rule $f^D \in \mathcal{F}^C$ minimizes the distortion if and only if for any object $x \in K$, we have

$$x \in f^D(P) \iff \begin{cases} \sum_{i\in N_x(P)} \underline{\varepsilon}_i^x & \geq \sum_{i\notin N_x(P)} \underline{\varepsilon}_i^x; \quad \text{or} \\ \sum_{i\in N_x(P)} \underline{\varepsilon}_i^x \sum_{i\in N_x(P)} \bar{\varepsilon}_i^x & \geq \sum_{i\notin N_x(P)} \underline{\varepsilon}_i^x \sum_{i\notin N_x(P)} \bar{\varepsilon}_i^x. \end{cases} \tag{4.26}$$

Define a committee profile $\mathcal{W}_x^D$ for each $x \in K$ according to equation (4.26). Then by construction, $f^D \in \arg\max_{f\in\mathcal{F}^C} V^D(P, f)$ for all $P \in \mathcal{S}^N$. ∎

The committee structure of distortion minimizing voting-by-committees rules is quite similar to that of minimax regret voting rules: the first term guarantees minimum distortion when there is no "clear" decision and the second term guarantees the selection of an object if such decision is optimal for all possible realizations of $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)_{i\in N} \in H(P)$. Both criterion is based on the worst-case analysis: they compare the social welfare of a voting rule to that of the social optimum and then, minimize the differences with respects to all possible realizations of utility profile. Note that $f^D \in \mathcal{F}^C$ is second best, in the sense that it is distortion minimizing among the family of voting-by-committees rules but not among the set of all voting rules. As the valuation function $V^D$ is nonlinear, the distortion of a voting rule is not equal to the sum of distortion caused by decision on each object. In fact, Caragiannis and Procaccia (2011) shows that there is a lower bound on the distortion of any deterministic voting rules which increases exponentially with the number of alternatives.

To conclude this section, we present the following corollaries discussing two special cases of our analysis. The first case is when the social planner faces the same amount of parameter uncertainty from each voter, that is $(\underline{\varepsilon}_i^x, \bar{\varepsilon}_i^x) \to (\underline{\varepsilon}^x, \bar{\varepsilon}^x)$ for all $i \in N$.

**Corollary 4.1** *As $(\underline{\varepsilon}_i^x, \bar{\varepsilon}_i^x) \to (\underline{\varepsilon}^x, \bar{\varepsilon}^x)$ for all $i \in N$, $f^H, f^R$ and $f^D$ converges to* *the simple majority rule, that is weighted majority rule with* $w_i^x = 1$ *for all* $i \in N$.

The second special case is when the social planner receive more and more information, to an extend that he can construct a complete preference relation for each individual voter, that is $\underline{\varepsilon}_i^x, \bar{\varepsilon}_i^x \rightarrow \varepsilon_x^i$.

**Corollary 4.2** *As* $\underline{\varepsilon}_i^x, \bar{\varepsilon}_i^x \rightarrow \varepsilon_i^x$ *for all* $i \in N$, $f^H, f^R$ *and* $f^D$ *converges to the weighted majority rule with* $w_i^x = \varepsilon_i^x$ *for all* $i \in N$.

## 4.6 Conclusion

We have extended the characterization result of Barberà et al. (1991) to the domain of incomplete and separable preferences. However, along the way, to accommodate the indecisiveness of agents, we impose a stronger notion of incentive compatibility, termed justifiable strategy-proofness. We show that the characterization of justifiable strategy-proof voting rules is similar to that of strategy-proof rules in a complete and separable domain. Furthermore, we also study the cardinal consequences of ordinal-based voting rules where social planner has strict uncertainty of agents' utility function. The optimal choice of voting rules depends on (i) the criterion on which the social planner evaluates his lack of information and, (ii) the information he manages to extract from individual agents.

Although justifiable strategy-proofness provides a nice characterization of incentive compatible rule, in many cases that independence is not applicable, the family of rules that satisfy the standard notion of strategy-proofness is not yet characterized. Furthermore, in this paper, we consider a simple dichotomous preference over each individual object. As we enlarge the domain to allow for multinary level of inputs and output, voting-by-committees might not even be justifiable strategy-proof as monotonicity is no longer necessary for incentive compatibility constraint.

# Bibliography


Alesina, A., and Angeletos, G. (2005a). Fairness and redistribution: US vs. Europe. *American Economic Review* 95, 913–935.

Alesina, A., and Angeletos, G. (2005b). Redistribution, corruption and fairness. *Journal of Monetary Economics*, 1227–1244.

Alesina, A., and Giuliano, P. (2010). Preferences for redistribution. In Benhabib, J., Jackson, M., Bisin, A., (Eds.), *Handbook of Social Economics*, Vol. 1A, North Holland, Amsterdam.

Alesina, A., Cozzi, G., and Mantovan, N. (2012). The evolution of ideology, fairness, and redistribution. *Economic Journal* 122 (565), 1244–1261.

Almås, I., Cappelen, A., Sørensen, E., and Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science* 328, 1176–1178.

Arrow, K. J. (1950). A difficulty in the concept of social welfare. Journal of political economy, 58(4), 328-346.

Aumann, R. J. (1962) Utility theory without the completeness axiom. *Econometrica* 30, 445-462.

Austen-Smith, D., and Banks, J. (1999). *Positive Political Theory I: Collective Preference*. Ann Arbor: University of Michigan Press.

Azrieli, Y., Kim, S. (2014). Pareto efficiency and weighted majority rules. *International Economic Review*, 55(4), 1067-1088.

Banks, J. S. (1995). Singularity theory and core existence in the spatial model. *Journal of Mathematical Economics*, 24(6), 523-536.

Banks, J. S., Duggan, J., and Le Breton, M. (2006). Social choice and electoral competition in the general spatial model. Journal of Economic Theory, 126(1), 194-234.

Barberà, S., Sonnenschein, H. and Zhou, L. (1991) Voting by committees. *Econometrica* 59, 595-609.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Bolton, G., and Ockenfels, A. (2000). ERC–a theory of equity, reciprocity, and competition. *American Economic Review* 90 (1), 166–93.

Boutilier, C., Caragiannis, I., Haber, C., Lu, T., Procaccia, A. D., and Sheffet, O. (2015) Optimal social choice functions: A utilitarian view. *Artificial Intelligence* 227, 190-213.

Boutilier, C., and Braziunas, D. (2007) Minimax regret based elicitation of generalized additive utilities. *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, 25-32.

Bräuninger, T. (2007). Stability in Spatial Voting Games with Restricted Preference Maximizing. *Journal of Theoretical Politics*, 19(2), 173-91.

Carlier, G., and Dana, R.A. (2013) Pareto optima and equilibria when preferences are incompletely known. *Journal of Economic Theory* 146(4), 1606-1623.

Caragiannis, I., and Procaccia, A. D. (2011) Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence* 175, 1655-1671.

Cho, W.J., and Ju, B.-G. (2017) Multinary group identification . *Theoretical Economics* 12, 513-531.

Carroll, G. (2010). An efficiency theorem for incompletely known preferences. *Journal of Economic Theory*, 145, 2463-2470.

Clark, A., and D'Ambrosio, C. (2015). Attitudes to income inequality: Experimental and survey evidence. In Atkinson, A., Bourguignon, F., (Eds.), *Handbook of Income Distribution*, Volume 2A, Chapter 13, Elsevier, 1147–1208.

Danan, E., Gajdos, T., Tallon, J. M. (2015). Harsanyi's aggregation theorem with incomplete preferences. *American Economic Journal: Microeconomics*, 7(1), 61-69.

Danan, E., Gajdos, T., Hill, B., Tallon, J-M. (2016). Robust Social Decisions. *American Economic Review*, 106(9), 2407-2425.

Dawes, C., Fowler, J., Johnson, T., McElreath, R., and Smirnov, O. (2007). Egalitarian motives in humans. *Nature* 446, 794–796.

Dawes, C., Loewen, P., Schreiber, D., Simmons, A., Flagan, T., McElreath, R., Bokemper, S., Fowler, J., and Paulus, M. (2012). Neural basis of egalitarian behavior. *Proceedings of the National Academy of Sciences* 109 (17), USA, 6479–6483.

Debowicz, D., Saporiti, A., and Wang, Y. (2017). On altruistic and electoral income redistribution: Theory and data. Available at https://sites.google.com/site/adsaporiti/home.

Dhami, S., and al-Nowaihi, A. (2010a). Redistributive policy with heterogenous social preferences of voters. *European Economic Review* 54 (6), 743–759.

Dhami, S., and al-Nowaihi, A. (2010b). Existence of a Condorcet winner when voters have other regarding preferences. *Journal of Public Economic Theory* 12 (5), 897–922.

Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press, Oxford. Chapter 6, 398–452.

Dimick, M., Rueda, D., and Stegmueller, D. (2017). The altruistic rich? Inequality and other-regarding preferences for redistribution. *Quarterly Journal of Political Science* 11 (4), 385–439.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper.

Dubra, J., Maccheroni, F., Ok, E. A. (2004). Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115(1), 118-133.

Duggan, J. (2018). Necessary gradient restrictions at the core of a voting rule. *Journal of Mathematical Economics*, 79,1-9.

Duggan, J. and Schwartz, T. (2000) Strategic manipulability without resoluteness or shared beliefs: Gibbard-Satterthwaite generalized. *Social Choice and Welfare* 17(1), 85-93.

Eguia, J. X. (2011). Foundations of spatial preferences. *Journal of Mathematical Economics*, 47(2), 200-205.

Eguia, J. X. (2013). On the spatial representation of preference profiles. *Economic Theory*, 52(1), 103-128.

Elster, J. (Ed.). (1987). The multiple self. Cambridge University Press.

Engelmann, D., and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94 (4), 857–869.

Engelmann, D., and Strobel, M. (2007). Preferences over income distributions. Experimental evidence. *Public Finance Review* 35, 285–310.

Fehr, E., and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817–868.

Fehr, E. (2009). Social preferences and the brain. In: *Neuroeconomics: Decision Making and the Brain*, Glimcher, P., Camerer, C., Poldrack, R., Rangel, A., Fehr, E. (Eds), Elsevier, 215–232.

Flamand, S. (2012). Heterogeneous social preferences in a model of voting on redistribution. Unpublished manuscript, Universidad Autónoma de Barcelona, 1-26.

Galasso, V. (2003). Redistribution and fairness: a note. *European Journal of Political Economy* 19 (4), 885–892.

Gershkov, A., Moldovanu, B., Shi, X. (2017). Optimal voting rules. *The Review of Economic Studies*, 84(2), 688-717.

Gibbard, A. (1973) Manipulation of voting schemes: a general result. *Econometrica* 41, 587-601.

Greenberg, J. (1979). Consistent Majority Rules over Compact Sets of Alternatives. *Econometrica*, 47(3), 627-636.

Hatsumi, K., Berga, D., and Serizawa, S. (2014), A maximal domain for strategy-proof and no-vetoer rules in the multi-object choice model. *International Journal of Game Theory* 43, 153-168.

Heller, Y. (2012). Justifiable choice. *Games and Economic Behavior* 76, 375-390.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, 39(153), 41-57.

Humphreys, M., and Laver, M. (2010). Spatial Models, Cognitive Metrics, and Majority Rule Equilibria. *British Journal of Political Science*, 40(1), 11-30.

Kasher, A., and Rubinstein, R. (1997). On the question "who is a J": A social choice approach. *Logique Analyse* 160, 385-395.

Ju, B.-G. (2003). A characterization of strategy-proof voting rules for separable weak orderings. *Social Choice and Welfare* 21, 469-499.

Ju, B.-G. (2005). An efficiency characterization of plurality social choice on simple preference domains. *Economic Theory* 26(1), 115-128.

Levy, G. (2004). A model of political parties. *Journal of Economic Theory*, 115(2), 250-277.

Lindbeck, A., and Weibull, J. (1987). Balanced-budget redistribution as the outcome of political competition. *Public Choice* 52 (3), 273–297.

Le Breton, M., and Sen, A. (1999). Separable preferences, strategyproofness, and decomposability. *Econometrica* 67(3), 605-628.

Luttens, R., and Valfort, M. (2012). Voting for redistribution under desert-sensitive altruism. *Scandinavian Journal of Economics* 114 (3), 881–907.

Matthews, S. (1980). Pairwise Symmetry Conditions for Voting Equilibria. *International Journal of Game Theory*, 9, 141-156.

McKelvey, R. D. (1976). Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic theory*, 12(3), 472-482.

McKelvey, R. D. (1986). Covering, dominance, and institution-free properties of social choice. *American Journal of Political Science*, 283-314.

McKelvey, R. D., and Schofield, N. (1987). Generalized Symmetry Conditions at a Core Point. *Econometrica*, 55(4), 923-933.

McKelvey, R. D., and Wendell, R. (1976). Voting Equilibria in Multidimensional Choice Spaces. *Mathematics of Operational Research*, 1(2), 144-158.

Meltzer, A. H., and Richard, S. F. (1981). A rational theory of the size of government. *Journal of Political Economy* 89, 914–927.

Miller, N. R., Grofman, B., and Feld, S. L. (1989). The geometry of majority rule. *Journal of Theoretical Politics*, 1(4), 379-406.

Nehring, K., and Puppe, C. (2007). The structure of strategy-proof social choice, Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory* 135(1), 269-305.

Ok, E. A. (2002) Utility Representation of an Incomplete Preference Relation. *Journal of Economic Theory*, 104(2), 429-449.

Plott, C. R. (1967). A Notion of Equilibrium and its Possibility Under Majority Rule. *The American Economic Review*, 57(4), 787-806.

Rilling, J., and Sanfey, A. (2011). The neuroscience of social decision-making. *Annual Review of Psychology* 62, 23–48.

Rae, D. W. (1969). Decision-rules and individual values in constitutional choice. *American Political Science Review*, 63(1), 40-56.

Rae, D., and Taylor, M. (1971). Decision Rules and Policy Outcomes. *British Journal of Political Science*, 1, 71-90.

Roberts, A. W., and Varberg, D. E. (1973). *Convex Functions*. Academic Press, New York.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press. Princeton, NJ.

Samet, D., and Schmeidler, D. (2003). Between liberalism and democracy. *Journal of Economic Theory* 110, 213-233.

Satterthwaite, M. (1975) Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions, *Journal of Economic Theory* 10, 187-217.

Saari, D. G. (1997). The generic existence of a core for q-rules. *Economic Theory*, 9(2), 219-260.

Schofield, N. (1978). Instability of simple dynamic games. *The Review of Economic Studies*, 45(3), 575-594.

Shepsle, K. A. (1979). Institutional arrangements and equilibrium in multidimensional voting models. *American Journal of Political Science*, 27-59.

Sloss, J. (1973). Stable Outcome in Majority Voting Games. *Public Choice*, 15, 19-48.

Smale, S. (1973). Global Analysis and Economics I: Pareto Optimum and a Generalization of Morse Theory. *Dynamical Systems*, New York: Academic Press.

Tabibnia, G., Satpute, A., and Lieberman, M. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science* 19, 339–347.

Tricomi, E., Rangel, A., Camerer, C., and O'Doherty, J. (2010). Neural evidence for inequality-averse social preferences. *Nature* 463 (7284), 1089–1091.

Tyran, J-R., and Sausgruber, R. (2006). A little fairness may induce a lot of redistribution in democracy, *European Economic Review* 50 (2), 469–485.

Wendell R., and Thorson, S. (1974). Some generalizations of social decisions under majority rule. *Econometrica*, 42, 893-912.

Zaki, J., and Mitchell, J. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences* 108 (29), USA, 19761–19766.