Social Choice and Logic via Simple Games

MSc Thesis (*Afstudeerscriptie*)

written by

Tijmen R. Daniëls
(born 18 March 1978 in Haarlem)

under the supervision of **Dr. Eric Pacuit**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:	**Members of the Thesis Committee:**
24th April 2007	Prof.dr. J.F.A.K. van Benthem
	Prof.dr. P. van Emde Boas
	Prof.dr. M.C.W. Janssen
	dr. E. Pacuit

Institute for Logic, Language and Computation

La savoir, du reste, n'est pas une qualité naturelle de l'homme. Bien qu'il admette une moyenne avec deux limites, et qu'on puisse supposer que, dans une nation, toutes les nuances intermédiaires sont parcourues conformément à la loi de possibilité, cependant on ne doit pas considérer cette moyenne comme le type de la perfection; elle ne fait que constater le degré de diffusion des lumières, et la limite supérieure indique le degré d'avancement absolu de la science; c'est tout ce qu'on en peut déduire.

—Adolphe Quételet (1796–1874), *Du système social et des lois qui le régissent.*

1 Introduction

1.1 The Wisdom of Crowds

This thesis is about social choice theory and logic. Social choice theory is concerned with questions of how a group of agents can decide as a collective in a way that reflects the individual opinions of those involved. The following example, known as the **doctrinal paradox** [23], highlights the link between social choice theory and logic.[1]

Example 1. *Suppose a group of three experts is asked to form an opinion on the following three propositions: p, q, and p* implies *q. After some consideration, the first expert declares that he subscribes to p* implies *q (but not to p or q), the second subscribes to solely to p and the third to all three propositions. Note that all opinions are individually rational, in the sense that they are logically coherent.*

Suppose proposition-wise majority voting is used to distill the 'collective opinion' of the three experts. There is a majority in favour of p, and a majority acknowledges that p implies *q, however there is no majority favouring q. In other words, under the majority rule, the experts' opinions are not "collectively rational"—the outcome of such a "judgement aggregation procedure" is logically inconsistent.*

Proposition-wise majority voting can be seen as a way to *transform*—or *aggregate*—individual opinions into what might be called a "collective opinion". The example suggests that under such transformations from the individual to the collective level, something happens to the logical rules that governs these opinions.

This thesis falls into two parts. The first part is concerned with the consequences of the observation that the logical consistency of individual opinions can be lost lost when passing to the collective level. This part investigates the constraints that are placed upon the process of collective decision making if it is to preserve logical consistency. We apply the methods of social choice theorists to the subject matter that concerns logicians.

[1] A similar problem is discussed in the computer science literature. See [38].

One can also adopt a slightly more lenient perspective, recognising that collective opinions are simply governed by other logical rules than individual opinions. In the second part of this thesis we enquire what such rules may look like. This investigation makes the logic of collective agency an object of study in its own right. In our investigation, we will draw links between the methodology of social choice theorists and that of logicians.

1.2 Social Choice Theory and Logic

Social choice theory has rich history that can be traced back more than two centuries, to eighteenth century thinkers Jeremy Bentham, Jean-Charles de Borda, and especially the Marquis de Condorcet. These mathematicians were concerned with the design and analysis of voting systems according to scientific principles; a theme that attracted the attention of many great minds in the age of enlightenment, infatuated with the dream of building society upon reason. The phrase "social choice theory" still reflects this ideal of applying formal methods to the realm of the social.

For many years the work done by these thinkers laid dormant. Then it suddenly picked up steam in the 1950's, when the mathematical economist Kenneth Arrow used observations originally made by Condorcet to prove a striking result, *viz.*, that it is impossible to aggregate rational preference relations into a collective (or social) rational preference relation by a mathematical procedure that satisfies certain natural—'democratic'—desiderata [2].[2] In other words, a collective of agents—a "society"—may lack well-defined preferences, even if the preferences of all the individual agents of the society are rational. Consequently, the idea that a decision or state of affairs might be desirable from viewpoint of an individual has no natural counterpart in terms of "social desirability".

Arrow's *impossibility theorem*, as it became known[3], foretold gloomy prospects for founding collective decision-making on microeconomic principles. One particular implication of the theorem is that one cannot find a natural measure for aggregate welfare that is grounded systematically in the individual preferences of the agents that make up an economy. Therefore it may be impossible to evaluate, say, the social welfare effects of a policy measure in terms of how it affects these individual agents, however natural and desirable this idea might seem *a priori*. The paradoxical nature of this result inspired a research programme aimed at investigating the mathematical boundaries to collective decision-making and social welfare rankings based on individuals' preferences. In the wake of Arrow's impossibility theorem, many startingly similar results followed. Sen [39] may be consulted for a broad introduction to the field.

[2] A rational preference relation is a binary, connected, reflexive, and transitive relation.

[3] Ironically, Arrow refers to his result as the 'general possibility theorem', but appears to be alone in this choice of terminology. For a lucid exposition of the proof, refer to Mas-Colell et al. [25].

As can be observed from the example above, problems of aggregation are not restricted to rational preference relations. In fact, the question of the existence of a social preference relation is but one of the issues surrounding collective decision making. Some recent work on social choice has revolved around **judgement aggregation** (a non-exhaustive list includes [8], [10], [14], [24], [32], [36]). Formally, this concerns the question of partitioning off a set of collectively accepted statements from a larger collection of interconnected logical sentences, in a logically consistent way and by a method reflecting the individual views of the agents as much as possible. In some sense the story of judgement aggregation appears as a case of history repeating. Judgement aggregation can superficially be regarded as a generalisation of preference aggregation—it is by now well established that many of the Arrovian-style impossibility results on preference aggregation have counterparts in this newer context. And indeed, the interest in judgement aggregation was spawned initially by the discovery of an impossibility theorem, by List and Petitt [24], in a paper that connects the doctrinal paradox to social choice theory. Dietrich and List [10] have shown that this result is in fact a close analogue of Arrow's theorem.

A close look at the way these two kinds of problems have been treated in the literature reveals resemblances that run more than skin deep. The Arrovian theorems of judgement aggregation further unclothe the ambit of Arrow's paradoxical result. In our view, there are some merits of judgement aggregation over preference aggregation that warrant the renewed interest. First, by investigating the boundaries of collective reasoning from a purely logical stance, judgement aggregation elevates the theory of social choice to a higher level of abstraction as well as to a broader, and perhaps more natural, conceptualisation of the "rationality of the collective" than is provided by the focus on preference relations stemming from economics. Second, judgement aggregation explicitly brings out the connection between logic and social choice theory.

Historically, social choice theory has adopted an "axiomatic" approach, and thus an implicit link with logic has always been present. It is interesting to note that Kenneth Arrow was introduced to the relevant axiomatic mathematical methods by Alfred Tarski [41]. However, a shift of focus from the aggregation of preference relations to logical statements turns up deeper connections with logic. As noted, collective decisions appear to answer to idiosyncratic logical rules. A promising way to study these rules is by defining a formal language that handles the peculiar logic of collective agency. Recently, Pauly [34] has proposed a modal-flavoured logic that does just that.

1.3 Simple Games

In this text we wish to highlight some connections between logic and social choice theory. If any general statement can be made about the way social choice theory has developed since Arrow's striking discovery, then it is that the literature has grown into a vast and daunting body of work. In this light it is all the more remarkable that many of the themes put forward in this introduction can be found in rudimentary form in in an important article by G. Th. Guilbaud [17], published in a 1952 special issue of the journal Économie Appliquée devoted entirely to Arrow's theorem (Arrow also contributed to this issue: see [1]). In this article, Guilbaud introduces the concept of *families of majorities*, touches upon general questions of preservation of mathematical structure under aggregation, and notes the consequences for logically interconnected propositions.[4] Guilbaud's article sparked a research programme conducted by a surrounding group of French mathematicians at the *Centre d'Analyse et de Mathématique Sociale*. Much of the group's research has been published in French, and as a consequence, some ideas contributed by the group have gone entirely unnoticed by other social choice theorists, and others have appeared only in disguised form. Bernard Monjardet, one of the most active members of this circle, has recently provided an historical and thematic overview of some their research [31].

A prominent theme in Monjardet's work, stressed for instance in [29], is the importance of simple games to understand results from social choice theory. Logicians typically swiftly recognise that familiar (to them) concepts such as filters and ultrafilters lie at the basis of impossibility results, and the theory of simple games provides an encompassing perspective that is able to accommodate such observations. For instance, the families of majorities alluded to above are also simple games. The notion of simple games, models for voting systems, stems from game theory, and is discussed extensively by Von Neumann and Morgenstern [43], and also by Shapley [40]. Arrow already noticed this interface between his own result and game theory; and the relation has been explored, amongst others, by Wilson [44].

1.4 Contributions of the Thesis

Our aim in this thesis will be to investigate the three sides of the triangle of logic, social choice, and game theory, through these simple games. To this end, section 2 introduces the theory of simple games in a self-contained way, yet with clear focus on the links with objects more familiar to logicians. Section 3 highlights parallels between some results on simple games and the newer literature on judgement aggregation. Most of the results presented in this section are generalisations of work due to Monjardet [29] on simple games and social choice theory, which are then applied in the new context. There are good reasons that justify closer scrutiny of the judgement aggregation impossibility results, such as the analogue of Arrow's theorem obtained by Dietrich and List [10], in this fashion. A basic motivation is that this approach helps to reveal the mathematical structure that leads up to impossibility results, and hence provides insight in how the judgement aggregation literature ties up with the older literature of social choice theory.

[4] In fact the result proved by List and Pettit [24] is informally already present in Guilbaud's article.

As an example, we present a "judgement aggregation" version of Maskin's celebrated theorem on "maximal transitivity" of the majority rule. Our proof makes use of a straightforward observation on the properties of a certain class of simple games, and because of this differs in flavour from that of Maskin, and immediately generalises to the judgement aggregation framework.

However, our intent is not simply to present new results in old bottles or vice versa. An approach with simple games at the basis transgresses the borders of social choice and naturally leads into logical territory, skirting results such as Łoś's theorem, and the theory of monotonic modal logics. Consequently, in section 4, we study **majority logic**, a variant of Pauly's logic of collective agency, which allows us to explicitly draw links between judgement aggregation and monotonic modal logic. We use tools from modal logic to study the expressivity of majority logic.

Since the application of modal methods to the business of social choice theory is unexplored territory, more or less all the results presented in section 4 are new, even though many proofs use arguments that are quite standard in the context of modal logic. The results share a common theme in their focus on the interaction between the logical strength of majority logic and the the axiomatic method of social choice theorists. However, a "language oriented approach" leads us to entertain different questions than usually found in social choice theory. Two main results deserve to be mentioned here. First, we show that certain kinds of axioms (namely the ones that can be represented in the language) cannot express non-dictatorship; in effect a very simple kind of "impossibility theorem". Second, we work towards simple characterisation results to show how this minimalistic logic fits into the picture of modal logic, which is a strictly stronger language. These two results fit in closely with the research agenda set forth by Pauly [35], who proposes to pay greater attention to the rôle of language in social choice theory. In essence, our results delineate how the majority logic sits in the hierarchy of logical strength.

Wrapping up, section 5 concludes the thesis with a bird's eye view of the relevance of the results that can be obtained by a focus on simple games.

2 Majorities and Simple Games

The focus of social choice theory is always on a collective of agents that has to decide as a group. Let N denote this set of agents; for much of what is said in this thesis N can be any nonempty finite or infinite set. Perhaps the most ubiquitous, familiar and natural decision rule to make group decisions is that of simple majority voting. In this section, we introduce simple games, models for voting power that provide a generalised interpretation of the notion of what is "majority".

Simple games are the primitive objects that underly most of the results that we will present later in this thesis. The objective of this section will be to introduce those properties of simple games that will play a rôle in the rest of this text in a coherent manner. In the later sections we shall investigate the correspondences between these simple games and other, more complex, group decision processes. However, the theory of simple games is itself a field of study with a remarkable history. Simple games were studied extensively by Von Neumann and Morgenstern [43], and the notion goes back at least to Dedekind. A recent book by Taylor and Zwicker [42] provides an excellent introduction to the general theory of simple games.[5]

2.1 Basic Definitions

Certainly, whatever notion of what is a "majority" one might adopt, if some subset A of the collective of agents, N, constitutes a majority of N, then any other subset B of N that properly contains A will also be a majority. This basic intuition may be formalised as follows. Let $W \subseteq \mathscr{P}(N)$ be the collection of subsets of N that we think of as the majorities of N. Then W is closed under supersets:

$$\text{if } A \in W \text{ and } A \subseteq B \subseteq N, \text{ then } B \in W. \tag{M1}$$

Definition 2. *Let N be a set of agents and $W \subseteq \mathscr{P}(N)$ be any collection that satisfies condition* (M1)*. The pair (N, W) is called a* ***simple game****.*

A simple game can be seen as a model for a voting system, where a certain coalition of agents can pass a proposal just in case it is an element of W. In game theoretic parlance, the elements of W are the **winning coalitions** of N. In this context, the monotonicity condition (M1) is natural. More generally however, W might be a collection that does not satisfy this condition, and in this case (N, W) isn't called a simple game but a **hypergraph**. In this text, we will not be concerned with such hypergraphs, though many of the concepts and results that we discuss can be generalised to this more general setting.

Two properties will play a fundamental rôle in the rest of this thesis. A simple game is called **proper** if it satisfies:

$$A \in W \text{ implies } N - A \notin W. \tag{M2a}$$

A simple game is called **strong** if it satisfies:

$$A \notin W \text{ implies } N - A \in W. \tag{M2b}$$

If W satisfies (M2a), then A is a majority of N only if its complement isn't; that is to

[5]Taylor and Zwicker are primarily concerned with the special class of "weighted" simple games—those games that can be derived from a weighting of the relative importance of the players.

say, all majorities are **strict**. On the other hand (M2b) expresses that A is a majority whenever its complement isn't; and thus in some sense that W has "enough" winning coalitions. G. Th. Guilbaud [17] called those simple games that are proper and strong **families of majorities**, and we will stick to this terminology below. In the literature on simple games, they are also often called **decisive** simple games.[6]

A game might contain players who are irrelevant in determining whether some set of agents is a majority. These players are called dummy players.

Definition 3. *A player is called a **dummy player** of* (N, W) *if:*

$$\text{for all } X \in \mathscr{P}(N),\ X \in W \iff X \cup \{i\}. \in W$$

Generalising this notion to sets, a set $A \subseteq N$ *is called a **set of dummy players** if:*

$$\text{for all } X \in \mathscr{P}(N),\ \text{and any } B \subseteq A,\ X \in W \iff X \cup B \in W.$$

Given $\Omega = (N, W)$*, denote the set of all dummy players by* $\mathscr{D}(\Omega)$*.*

Assuming the axiom of choice (and we will freely do so in this thesis), A is a set of dummy players if and only if each $i \in A$ is a dummy player.

Consider a game $\Omega = (N, W)$, and a set of players $A \dot{\cup} N$ (here, the symbol $\dot{\cup}$ denotes disjoint union, and thus carries the connotation that $N \cap A = \varnothing$). Since dummy players are irrelevant, one can easily obtain a game Ω' with the set A added to Ω as dummy players:

$$\Omega' = (N \dot{\cup} A, W') \text{ where } W' := \{X \in \mathscr{P}(N \dot{\cup} A) \mid X \cap N \in W\}.$$

Definition 4. *Let* $\Omega = (N, W)$ *be a simple game. A game* Ω' *that is obtained by adding a set of dummy players* A *to* N *is called a **dummy extension** of* Ω*. If* $\mathscr{D}(\Omega) \neq \varnothing$*, a game* Ω' *that is obtained by adding a set of dummy players* A *to* N *is also called a **dummy expansion** of* Ω*. If* $B \subseteq \mathscr{D}(\Omega)$*, then the game* $\Omega' := (N - B, W \cap \mathscr{P}(N - B))$ *is called a **dummy contraction** of* Ω*.*

Our definition of a simple game leaves room for two embarrassing systems that consist entirely of dummy players. The first kind are the **verum** games of the form $(N, \mathscr{P}(N))$, in which every subset of players constitutes a majority. The second kind are the **trivial** games of the form $(N, \varnothing)$, in which no collection of players is a majority. Note that verum games are strong but not proper, and that the converse is true for trivial games. If a simple game is trivial or verum, we will say it is **degenerate**.

[6] The simple games studied by Von Neumann and Morgernstern are both proper and strong. They investigated various properties of such games, including issues of computational complexity.

Since the collection of winning coalitions of a simple game is closed under taking supersets, a full specification of all the winning coalitions of a simple game typically introduces a large degree of redundancy.

Definition 5. *Let (N, W) be a simple game. The* ***non-monotonic core*** *of (N, W) is defined by:*

$$\mathscr{C}((N, W)) := \{A \in W \mid \textit{there is no } B \in W \textit{ s.t. } B \subset A\},$$

that is, it is the collection of minimal winning coalitions of (N, W). A simple game (N, W) is said to ***contain its non-monotonic core*** *if its collection of winning coalitions can be generated from its core: if $W = \{A \in \mathscr{P}(N) \mid A \supseteq B \textit{ for some } B \in \mathscr{C}((N, W))\}$.*

As simple games are closed under taking supersets, a simple game that contains its non-monotonic core is completely determined by it. Clearly, every simple game over finite N contains its non-monotonic core, but this isn't true for some simple games over infinite N.[7]

2.2 Important Types of Simple Games

Given the ubiquity of simple majority voting, perhaps the most natural simple games are those that encode the **majority rule**, under which a group A is a winning coalition if it contains at least half of all agents in N.

Definition 6. *Let N be any finite set. The* ***Simple Majority Game*** *over N is the game (N, W) where W the subset of $\mathscr{P}(N)$ such that $A \in W$ if and only if $|A| \geq \frac{1}{2}|N|$.*

We now introduce some more general kinds of simple games, and discuss some ways in which they are related.

Families of Majorities and the Ultra-Property

G. Th. Guilbaud [17] studied systems (N, W) that satisfied (M1), (M2a), (M2b), and:

$$A \in W \text{ if and only if for all } B \in W, A \cap B \neq \varnothing, \tag{M3}$$

In words, this last condition says that A is a majority if and only if A intersects every other majority. However, this condition isn't independent of (M1), (M2a) and (M2b).

[7] For instance, it fails for all non-principal ultrafilters, introduced below; for such simple games the non-monotonic core is empty.

In fact, denote:

$$A \in W \text{ only if for all } B \in W, A \cap B \neq \varnothing, \tag{M3a}$$

$$A \in W \text{ whenever for all } B \in W, A \cap B \neq \varnothing, \tag{M3b}$$

Then one can show:

Lemma 7. *(a)* (M1) + (M2a) $\implies$ (M3a)*, and* (M3a) $\implies$ (M2a)*. Hence a simple game satisfies* (M3a) *if and only if it satisfies* (M2a)*;*
(b) (M2b) $\implies$ (M3b)*, and* (M3b) + (M1) $\implies$ (M2b)*. Hence a simple game satisfies* (M3b) *if and only if it satisfies* (M2b)*;*
(c) An arbitrary collection $W \subseteq \mathscr{P}(N)$ satisfies (M3) *if and only if it satisfies* (M1)*,* (M2a) *and* (M2b)*.*

Proof. (a $\Rightarrow$). Suppose (M1) and (M2a). Suppose $A \in W$. Now take some arbitrary $B \in W$. Then $A \cap B \neq \varnothing$—for suppose otherwise, then $a \in A$ implies $a \notin B$. It follows $N - B \supseteq A$. By (M1), $N - B \in \Omega$ and yet by (M2a), $N - B \notin W$: a contradiction. Conclude W satisfies (M3a).
(a $\Leftarrow$). Assume (M3a). Suppose $A \in W$. Then $(N - A) \cap A = \varnothing$, whence by (M3a) $N - A \notin W$, giving (M2a).
(b $\Rightarrow$). Assume (M2b) and suppose A intersects every $B \in W$. We will prove $A \in W$, and thus that (M3b) holds. Suppose, to the contrary $A \notin W$. Then by (M2b) we find $N - A \in W$. But $A \cap (N - A) = \varnothing$, contradicting our assumption on A.
(b $\Leftarrow$). Suppose (M3b) and (M1). Suppose $A \notin W$. Then there is $B \in W$ such that $A \cap B = \varnothing$—otherwise by (M3b), $A \in W$—and hence $b \in B$ implies $b \notin A$. But this means $B \subseteq N - A$, and by (M1), $N - A \in W$. Conclude (M2b).
(c) It remains to be shown that (M3) implies (M1). Let $C \supseteq A$. By (M3a), for all $B \in W$, $A \cap B \neq \varnothing$. But since $C \supseteq A$, then $C \cap B \neq \varnothing$ for all $B \in W$. By (M3b), $C \in W$. ■

G. Th. Guilbaud [17] called systems (N, W) that satisfied (M1), (M2a), (M2b), and (M3) **families of majorities**, and by lemma 7, these are precisely the simple games that are both proper and strong, in line with what we have suggested earlier. Condition (M3) gives another characterisation of such games. A familiar example of a family of majorities is the simple majority game when $|N|$ is odd. The following is a less familiar example, that will come in handy later on.

Example 8. ***The dualton game****. Let N be any set containing i, j, k. Let W be the smallest set containing $\{i, j\}$, $\{j, k\}$ and $\{i, k\}$ that is closed under supersets. Then $\Omega_D^N = (N, W)$ is a family of majorities.*

Proof. (M1) is clear. (M3a) is satisfied: if $A, B \in W$, then A contains, say $\{i, j\}$ as a subset, and B, say $\{i, k\}$, so $A \cap B \neq \varnothing$. Finally (M2b) is satisfied: if $A \notin W$, then it cannot contain both i and j. If it contains neither, then its complement contains both, so and $N - A \in W$. If, on the other hand, A contains, say i but not j, then either $k \in A$, and hence $\{i, k\} \subseteq A$; or $k \notin A$ and hence $\{j, k\} \subseteq N - A$; however the former contradicts our assumption $A \notin W$, and so the latter is surely the case. This gives $N - A \in W$. ■

The families of majorities are the simple games satisfy what might be called the **ultra-property**: for any set $A \subseteq N$, either A or its complement $N - A$ is a winning coalition. As will become more than apparent below, this property is important in logic. However, for modelling certain notions of "majority", families of majorities may sometimes be too strong. For instance, consider the finite set $N = \{1, 2, 3, 4\}$. If the opinions of all agents carry equal weight, one would like to label $\{1, 2\}$ a majority if and only if one calls $\{3, 4\}$ a majority—but the ultra-property precludes this. Other concepts, that capture other intuitions of what it means to constitute a majority, have been proposed in the literature.

Majority Spaces and Finite Anonimity

A permutation of N is a bijection $\sigma : N \to N$. Say that a permutation σ is **finite** if $\{i \in N \mid \sigma(i) \neq i\}$ is a finite set. A collection $W \subseteq \mathscr{P}(N)$ is called **closed under finite permutations**, if for every finite permutation $\sigma : N \to N$, $A \in W$ if and only if $\sigma[A] \in W$. Clearly, the Simple Majority Game is closed under finite permutations. In fact, if we order the simple games over N by the $\subseteq$ relation on their collections of winning coalitions, the Simple Majority Game is *minimal* among the finite strong simple games over N satisfying closure under finite permutations. This observation is easy and useful, and it is remarkable that we did not find any explicit reference to it in the social choice literature.

Lemma 9. *Let N be finite and $\Omega = (N, W)$ a strong simple game closed under (finite) permutations. Let $\Omega_S^N = (N, W')$ be the Simple Majority Game over N. Then $W' \subseteq W$.*

Proof. Since N is finite both Ω and Ω_S^N contain their non-monotonic cores. For $|N|$ is odd (even) the non-monotonic core of Ω_S^N contains exactly the subsets of N of cardinality $|N|/2 + 1$ $(N/2)$. Let $A \in \mathscr{C}(\Omega)$ and $B \in \mathscr{C}(\Omega_S^N)$; our aim is to show that $|A| \leq |B|$. As W is closed under finite permutations we may then conclude $W' \subseteq W$.

(case $|N|$ odd) We will prove $|A| \leq |N|/2 + 1$. Suppose not, and take $B \in \mathscr{P}(N)$ s.t. $|B| = |N|/2 + 1$. B is not a majority (after all, there is a permutation taking B to a proper subset of A, so $B \in W$ contradicts that $A \in \mathscr{C}(\Omega)$), so $N - B$ is. But $|N - B| = |N| - |B| = |N| - |N|/2 - 1 < |N|/2 + 1$. Hence, there exists a permutation carrying $N - B$ to a proper subset of $|A|$. This contradicts that $A \in \mathscr{C}(\Omega)$.

(case $|N|$ even) We will prove $|A| \leq |N|/2$. Suppose not. Then take $B \in \mathscr{P}(N)$ s.t. $|B| = |N|/2$. By the same argument as before, B isn't a majority, so $N - B$ is. But

$|N-B| = |B|$, so there exists a finite permutation taking $|N-B|$ to B. It follows B is a majority after all, but this is absurd. ■

Majority spaces are proposed by Eric Pacuit and Samer Salame in [33] as a generalisation of the simple majority game over a possibly infinite set of agents. A **majority space** is a strong simple game (N, W) that satisfies two additional conditions:

$$A \in W \text{ only if, for all } B \in W, \text{either } A \cap B \neq \varnothing \text{ or } B = N - A; \tag{M4}$$

$$W \text{ is closed under finite permutations.} \tag{M5}$$

The first condition says that if A and B are majorities such that $A \cap B = \varnothing$, then A and B are each others complements. The latter condition expresses what might be called finite anonymity. The Simple Majority Game, defined for finite sets N, satisfies both of these conditions, and indeed if N is a finite set the Simple Majority Game over N is in fact the only simple game over N that does. Moreover, it is quite easy to see that if Ω is a family of majorities, then Ω is also a majority space if and only if its underlying collection of winning coalitions is closed under finite permutations.

Given a simple game Ω, call a set $A \in W$ a **weak majority** if and only if $N - A \in W$, and the (unordered) set $\{A, N-A\}$ a **pair of weak majorities**. In contrast with families of majorities, the conditions on majority spaces allow for weak majorities. It is clear from (M4) that a majority space is proper (hence has the ultra-property) if it contains no weak majorities. This observation gives us a simple procedure for turning a majority space into a family of majorities, by simply removing one of any pair of weak majorities (in general, after such surgery Ω is no longer a majority space). *A fortiori,* if a majority space is already proper, it is also a family of majorities. The converse is not true, because the finite anonymity condition generally fails for families of majorities.

Ultrafilters, and Dictatorships versus Finite Anonimity

Let $\Omega = (N, W)$ be a simple game. If the intersection of every finite collection of majorities of W is itself again a member of W, then Ω is called a **filter**. (Our terminology comprises some *abuse de langage*: where we assign the name 'filter' to the simple game Ω, in set theory and topology the set W itself is called a filter). It is easy to show that a filter (N, W) is a proper simple game if and only if $\varnothing \notin W$.

A simple game (N, W) is called an **ultrafilter** if it is a *maximal* proper filter (that is, it cannot be expanded to another proper filter (N, W') by adding further winning coalitions to W). It is again easy to show that a simple game is an ultrafilter if and only if it is both a filter and a family of majorities. An ultrafilter is called **principal** if W contains exactly one singleton set and **non-principal** otherwise. If N is finite, then all the ultrafilters over N are principal. By monotonicity, the majorities of a principal ultrafilter are all the subsets containing some particular agent i of N; and hence a set of agents A is a winning coalition if and only if agent i is present in it. In plain language, the agent i is a dictator.

Definition 10. *Let N by a set of agents and fix some $i \in N$. The game (N, W), with $W = \{A \in \mathscr{P}(N) \mid i \in A\}$, is called a* ***dictatorship****.*

Interestingly, non-principal ultrafilters are in some sense at the other end of the extreme: the following result shows that rather than concentrating all the power in the hands of one individual, non-principal ultrafilters satisfy the finite neutrality condition demanded by majority spaces. This observation ramifies the close connection between families of majorities and majority spaces that are closed under intersections.

Lemma 11. *Every non-principal ultrafilter is closed under finite permutations.*

Proof. It is facile to establish (and well known) that if $\Omega = (N, W)$ is an ultrafilter then (i) W contains only infinite subsets of N iff Ω is non-principal and (ii) that for any set $A \in W$, if $B \subseteq A$, then either $A - B \in W$ or $B \in W$. If σ is a finite permutation, the set $X = \{n \in N : \sigma(n) \neq n\}$ is finite. So if $A \in W$, then $A - X \in W$ by (i) and (ii). As ultrafilters are closed upwards, for any $Y \subseteq X$, $(A - X) \cup \sigma[Y] \in W$, in particular when $Y = X \cap A$. ∎

As an immediate corollary every non-principal ultrafilter is a majority space. This observation calls for some qualification. If (N, W) is an ultrafilter, and $A \in W$, then every bipartition of A again contains exactly one winning coalition. So while such a simple game does not contain a dictator, every winning coalition A can be contracted into a new winning coalition that is a strict subset of it. This process can be continued for ever and ever. In the literature on social choice theory the terminology *invisible dictatorship* has been coined for these rather unintuitive properties of non-principal ultrafilters (See Kirman and Sondermann, [22]).

2.3 Maximality Conditions and Characterisations

In the previous section, the principal ultrafilters have been characterised as the games that correspond to dictatorships. This exhausts the possible forms an ultrafilter may take on when N is finite. If N is infinite, then the ultrafilter lemma (see e.g. Hodges [19, Lemma 8.5.5] for a proof) guarantees a large supply of non-principal ultrafilters. A family of sets $\mathcal{F}$ has the **finite intersection property (FIP)** if the intersection of every finite subset of $\mathcal{F}$ isn't empty. Clearly the collection of winning coalitions of every filter has the FIP.

Lemma 12 (The Ultrafilter Lemma)**.** *Let $\mathcal{F}$ be a family of sets that have the finite intersection property. Let $N \supseteq \bigcup \mathcal{F}$. Then $\mathcal{F}$ can be expanded to an appropriate family $\mathcal{F}^*$ to obtain an ultrafilter $(N, \mathcal{F}^*)$.*

The ultra-property forces maximality upon proper simple games: if (N, W) satisfies the ultraproperty, then there is no proper simple game (N, W') such that $W' \supset W$. To see this, suppose $W' \supset W$. There is $A \in W'$ such that $A \notin W$. However, by the ultra-property implies $N - A \in W$. Hence both $A \in W$ and $N - A \in W$, and thus (N, W') isn't proper. By lemma 7, any simple game that satisfies the ultraproperty is a family of majorities, and thus these are exactly the maximal proper simple games. Ultrafilters, then, are the maximal proper simple games that are closed under finite intersection. One of the themes of this text is that families of majorities are an interesting generalisation of the notion of ultrafilters. Interestingly—though perhaps not too surprisingly—one can show an analogue to the ultrafilter lemma. A family of sets $\mathcal{F}$ has the **pairwise intersection property (PIP)** if no pairwise intersection of sets of $\mathcal{F}$ yields the empty set—this is the same as saying that $\mathcal{F}$ satisfies (M3a). Thus the collection of winning coalitions of every proper simple game has the PIP.

Lemma 13. *Let $\mathcal{F}$ be a family of sets that have the pairwise intersection property. Let $N \supseteq \bigcup \mathcal{F}$. Then $\mathcal{F}$ can be expanded to an appropriate family $\mathcal{F}^*$ to obtain a family of majorities $(N, \mathcal{F}^*)$.*

In fact, this observation is easy and useful, but we have not found its statement explicitly in the literature on simple games.

Proof. Let N be a set of agents. We will show first that if $\mathcal{C}$ is a $\subseteq$-chain of families of subsets of N each of which satisfies the PIP, then $\overline{\mathcal{F}} := \bigcup \mathcal{C}$ satisfies the PIP. Take $X, Y \in \overline{\mathcal{F}}$, we need to show that $X \cap Y \neq \varnothing$. Let $X \in \mathcal{F}' \in \mathcal{C}$ and $Y \in \mathcal{F}'' \in \mathcal{C}$. Without loss of generality we may suppose that $\mathcal{F}' \subseteq \mathcal{F}''$. Then $X \in \mathcal{F}''$. Now $\mathcal{F}''$ has the PIP, so $X \cap Y \neq \varnothing$.

Let $\mathcal{F}$ have the PIP. We use Zorn's Lemma to show that $\mathcal{F} \cup \{N\}$—which clearly also has the PIP—can be expanded to satisfy both (M3a) and (M3b). Let $(\mathcal{E}, \subseteq)$ be the ordered set of all possible PIP compatible extensions of $\mathcal{F} \cup \{N\}$, ordered by inclusion. By Zorn's Lemma and the observation above, each chain in $\mathcal{E}$ has a maximal element. Take a chain of families satisfying (M3a) and let $\mathcal{F}^*$ be maximal. We will show that for each set $A \subseteq N$, either $A \in \mathcal{F}^*$ or $N - A \in \mathcal{F}^*$. Clearly a family that has the PIP cannot contain both—so suppose $\mathcal{F}^*$ contains neither. Then by maximality there are sets B and C in $\mathcal{F}^*$ such that $A \cap B = \varnothing = (N - A) \cap C$. As $\mathcal{F}^*$ has the PIP, $D = B \cap C \neq \varnothing$. Then since $D \subseteq B$, $A \cap D = \varnothing$. Similarly, $N - A \cap D = \varnothing$. But then $A \cup N - A = N \cap D = \varnothing$ and so D is empty, a contradiction. Conclude that $\mathcal{F}^*$ satisfies (M2b). By lemma 7, (M2b) implies (M3b); together with (M3a) this suffices to show that $(N, \mathcal{F}^*)$ is a family of majorities. ■

Every maximal proper simple game is also a strong simple game. Furthermore, our discussion shows that the intersection properties of a family of sets are amongst the most primitive properties that determine the structure of maximal proper simple games that can be derived from such a set. Monjardet [30] observes that these two facts together can be used to provide a simple characterisation of the maximal proper simple games relative to the strong simple games. These are summed up in the following result.

Lemma 14. *Let (N, W) be a strong simple game.*
(a) (N, W) is a family of majorities if and only if for all $A \in W$ and $B \in W$, $A \cap B \neq \varnothing$.
(b) (N, W) is an ultrafilter if and only if for all sets $A, B, C \in W$, $A \cap B \cap C \neq \varnothing$.

Proof. The only non-obvious statement is (b, $\Leftarrow$). Let $\Omega = (N, W)$ be any simple game with $A \cap B \cap C \neq \varnothing$ for all $A, B, C \in W$. This implies in particular that $\varnothing \notin W$. Suppose $A, B \in W$. By assumption we have $(A \cap B) \cap C \neq \varnothing$ for all $C \in W$. By (M3b), $(A \cap B) \in W$, and thus W is closed under intersections. By induction, W has the FIP; thus Ω is a filter. Moreover, since $\varnothing \notin W$, Ω is a proper filter: let $A \in W$; $A \cap (N - A) = \varnothing \notin W$, and hence it follows from the fact that Ω is closed under intersections that $N - A \notin W$. So Ω satisfies (M2a). This means that Ω satisfies the ultraproperty, and thus is an ultrafilter. ■

2.4 Relations between Simple Games

Since the collection of winning coalitions of a simple game is a set of sets, simple games come equipped naturally with the inclusion ordering. That is, for simple games $\Omega = (N, W)$ and $\Omega' = (N', W')$, if $W \subseteq W'$, then every coalition that is winning in W is also winning in W'. We have already remarked (lemma 9) that the Simple Majority Game Ω_S^N is minimal for $\subseteq$ amongst the finite, 'anonymous', and strong simple games.

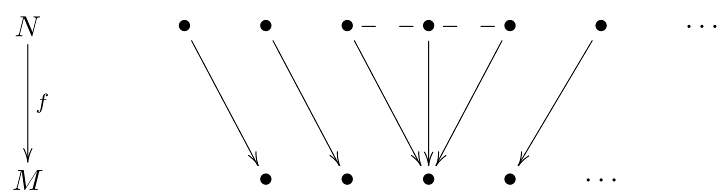
Our attention will be devoted to another hierarchy of interest. This is the **Rudin-Keisler ordering**, $\leq_{\text{RK}}$, introduced by Mary Ellen Rudin as an ordering of ultrafilters (see Jech [21]). Taylor and Zwicker [42] observe that this ordering has a natural interpretation when applied to simple games; and they point out this observation goes back to at least Isbell [20]. Let N and M be two sets of agents, and W a collection of winning coalitions of N. Suppose f is a map from N to M, then $f_*(W)$ is the subset of $\mathscr{P}(M)$ given by:

$$A \in f_*(W) \iff f^{-1}[A] \in W,$$

where $f^{-1}[A]$ is the pre-image of A, given by: $\{i \in N \mid f(i) \in A\}$.

In terms of simple games, the set $f_*(W)$ is obtained intuitively by considering W and imagining the players identified by f vote *en bloc*. Any coalition that is winning according to $f_*(W)$ can be formed as a winning coalition of (N, W) by block formation; see figure 1.

Definition 15. *We say that (N', W') is **RK-below** (N, W), if and only if there exists a map f such that $W' = f_*(W)$. In this case we write $(N', W') \leq_{\text{RK}} (N, W)$. We will write $(N', W') \leq_{\text{RK}}^{\text{S}} (N, W)$ if there exists a surjection f such that $W' = f_*(W)$.*



$f : N \to M$ identifies the three indicated points.

Figure 1: RK-projection

The above definition allows us to consider two relations, $\leq_{\text{RK}}$ and $\leq_{\text{RK}}^{\text{S}}$, over the class of simple games. Clearly $\leq_{\text{RK}}^{\text{S}} \subset \leq_{\text{RK}}$, and it isn't hard to see that $\leq_{\text{RK}}$ and $\leq_{\text{RK}}^{\text{S}}$ are transitive and reflexive relations ("pre-orderings").

Our definition of the Rudin-Keisler ordering differs from the one familiar from the literature on ultrafilters (and from the one given by Taylor and Zwicker [42]) in that we do not require f to be a surjection; in the literature on simple games, only the relation $\leq_{\text{RK}}^{\text{S}}$ is usually considered.[8] However, in this text, we choose $\leq_{\text{RK}}$ as the default. If $\Omega \leq_{\text{RK}} \Omega'$ then Ω is called an RK-projection of Ω'. Amongst other things, RK-projection preserves monotony—hence we can be confident that the projection of a simple game is again a simple game)—, and strongness and properness.

Two simple games Ω and Ω' are called **isomorphic** if $\Omega \leq_{\text{RK}} \Omega' \leq_{\text{RK}} \Omega$. The relation $\leq_{\text{RK}}$ partially orders the simple games up to isomorphism. Moreover:

Lemma 16. *If $(N \dot\cup A, W')$ is a dummy extension of (N, W), then $(N \dot\cup A, W')$ is isomorphic to (N, W).*

Proof. Since $(N \dot\cup A, W')$ is a dummy extension of (N, W), we have that $X \in W'$ if and only $X \cap N \in W$.

$(N \dot\cup A, W') \leq_{\text{RK}} (N, W)$: Consider the identity map from N to $N \dot\cup A$. Let $X \subseteq N \dot\cup A$. Then $\text{id}[W] = X - A$. Since A is a set of dummy players of $(N \dot\cup A, W')$, X is a winning coalition of $(N \dot\cup A, W')$ if and only if $X - A$ is winning in $(N \dot\cup A, W')$ (since A is a set of dummy players) if and only if $X - A = X \cap N = \text{id}[W]$ is winning in (N, W).

$(N, W) \leq_{\text{RK}} (N \dot\cup A, W')$: Pick $j \in N$ and let $f : N \dot\cup A \to N$ be defined by $f(i) = i$ if $i \in N$ and $f(i) = j$ if $i \in A$.

Let $X \subseteq N$. Then $X \cap N = X$. If $j \notin X$, then $f[X] = X$. In this case have that $X \in W$ if and only if $X = f[X] \in W'$. If $j \in X$, then $f[X] = X \cup A$. In this case we have $X \in W$ if and only if and only if $X \in W'$ if and only if $X \cup A = f[X] \in W'$ (since A is a set of dummy players). ■

[8] Taylor and Zwicker [42] do point out the possibility of dropping the surjectivity condition on f in this context.

The analogous claim for $\leq^{\mathrm{S}}_{\mathrm{RK}}$ quite obviously fails, which justifies our choice of $\leq_{\mathrm{RK}}$ as the default. In fact, it is quite easy to see that if $(N', W') \leq_{\mathrm{RK}} (N, W)$ and the projection function $f : N \to N'$ isn't surjective, then the set $N' - \mathrm{ran}(f)$ will consist of dummy players. Hence any RK-projection may be decomposed in a surjective projection and an operation that adds dummies.

A systematic investigation of this hierarchy is far beyond the scope of this text. However, in order to give a glimpse of its structure, we will prove two simple facts. One, on the subclass of non-degenerate simple games, the dictatorships are minimal for $\leq_{\mathrm{RK}}$.

Lemma 17. *Let Ω be any non-degenerate simple game and Ω' be a dictatorship. Then $\Omega' \leq_{\mathrm{RK}} \Omega$.*

Proof. Let $\Omega = (N, W)$ and $\Omega' = (N', W')$. Ω' is a dictatorship, hence there is i such that $A \in W'$ if and only if $i \in A$. Consider $f : N \to \{i\}$. On the one hand $X \in W'$ implies $i \in X$, and hence $f^{-1}[X] = N$; since Ω is non-degenerate, we know $W \neq \varnothing$, and hence $N \in W$ by monotony. On the other hand $X \notin \Omega'$ implies $i \notin X$, and hence $f^{-1}[X] = \varnothing$, and since Ω isn't degenerate, $W \neq \mathscr{P}(N)$, and hence (by monotony) $\varnothing \notin W$. ■

Two, there exist simple games that are incomparable under the Rudin-Keisler ordering.

Lemma 18. *Let $|N|, |M| \geq 3$. Let Ω^N_D be the dualton game of example 8 and consider the consensus game $\Omega^M_C := (M, \{M\})$. Then $\Omega^N_D \not\leq_{\mathrm{RK}} \Omega^M_C \not\leq_{\mathrm{RK}} \Omega^N_D$.*

Proof. ($\Omega^N_D \not\leq_{\mathrm{RK}} \Omega^M_C$). Let $f : M \to N$ be any function with the required properties; we will derive a contradiction. Since $\{i, j\}$ is a winning coalition of Ω^N_D, it must be that $f^{-1}[\{i, j\}]$ is a winning coalition of Ω^M_C, but this just means that $f^{-1}[\{i, j\}] = M$. Similarly, since $\{i, k\}$ is a winning coalition of Ω^N_D, it must be that $f^{-1}[\{i, k\}] = M$. But $f^{-1}[\{i, j\}] = M$ implies no $x \in M$ is mapped to k. Hence $f^{-1}[\{i\}] = M$. But then $f^{-1}[\{j, k\}] = \varnothing$ and $\varnothing$ isn't a winning coalition of Ω^M_C. However, $\{j, k\}$ is a winning coalition of Ω^N_c; a contradiction.
($\Omega^M_C \not\leq_{\mathrm{RK}} \Omega^N_D$). Suppose the contrary; it suffices to observe that the dualton game is strong, while the consensus game isn't. But strongness is preserved by RK-projection, giving a contradiction. ■

3 Judgement Aggregation and Arrovian Theorems

In this part we turn to the question of the existence of a "rational social opinion": whether a group or society of agents can form a logically coherent opinion on a collection of propositions, somehow taking into account the individual opinions of those involved. To investigate this question, we introduce a formal apparatus: the judgement aggregation framework. In this framework, a set of agents, N, has to decide collectively on a collection, T, of logical sentences uttered in a language, $\mathscr{L}$, according to some kind of social aggregation procedure. A social aggregation procedure is thus a method to single out a subset S of socially accepted sentences from T.

The first subsection introduces the judgement aggregation framework. In the subsection that follows it, we relate the framework to the problem of aggregating linear orderings, that has historically been much studied by social choice theorists. Thereafter, we turn to two kinds of relationships between the judgement aggregation framework and simple games. In subsection 3.3 we study a correspondence between the simple games and judgement aggregation procedures. In subsection 3.4, we study a duality result between families of majorities on the one hand, and domain restrictions on the set of logical sentences under collective scrutiny on the other. The final subsection of this part of the thesis combines the two ideas to elucidate some of the impossibility results that have been obtained by social choice theorists.

3.1 The Judgement Aggregation Framework

We assume $\mathscr{L}$ is a first order language.[9] A set of sentences in the language $\mathscr{L}$ is called a **theory**. A theory is **consistent** if it has a (first order) model, and **inconsistent** otherwise.[10] A theory is called **minimally inconsistent** if each of its proper subsets is consistent.

A sentence in the language $\mathscr{L}$ is a **negation** if it is of the form $\neg\alpha$. A sentence is called **positive** if it is either of the form β or of the form $[\neg\neg]\beta$—where β is not a negation and $[\neg\neg]$ is a string consisting of an even number of $\neg$'s. A first order theory $T \subseteq \mathscr{L}$ is a called an **agenda** if it satisfies the property that for each positive sentence α of $\mathscr{L}$, $\alpha \in T$ if and only if $\neg\alpha \in T$. A subset S of an agenda T will be called **complete** if and only if it contains exactly one of each pair $\alpha, \neg\alpha$ found in T. For the balance of this section we assume, without too much loss of generality, that no $\alpha \in T$ is a tautology or contradiction—clearly such sentences do not have to be voted upon.

Example 19. *Suppose agents have to decide on the statements $\alpha, \alpha \to \beta, \beta$. The set $\{\alpha, \neg\alpha, \beta, \neg\beta, \alpha \to \beta, \neg(\alpha \to \beta)\}$ is an agenda. If α and β are logically independent, the set $\{\alpha, \alpha \to \beta, \neg\beta\}$ is complete and minimally inconsistent.*

Let N stand for the set of agents; we imagine that each agent forms an individual opinion on the sentences in T. An **individual judgement set** is a complete subset of T. A **choice function** is a function $\pi : N \to \mathscr{P}(T)$, such that for each $i \in N$, $\pi(i)$ is the individual judgement set associated with the agent i. Hence such a choice function provides the individual opinions of agents; and consequently if a sentence α is

[9] We might as well identify a language with all its well-formed formulae (w.f.f.); when we write $S \subseteq \mathscr{L}$, we mean that S is a set of w.f.f. in the language $\mathscr{L}$.

[10] See e.g. Hodges [19] for first order model theory. We assume the logic is classical, that is, α is equivalent to $\neg\neg\alpha$.

an element of the individual judgement set $\pi(i)$, we say that agent i **accepts** α. Given a choice function π, we write $[\![\alpha]\!]_\pi$ for the set of agents that accept the sentence α, that is:

$$[\![\alpha]\!]_\pi := \{i \in N \mid \alpha \in \pi(i)\}.$$

For a set $S \subseteq T$ we will sometimes write $[\![S]\!]_\pi = A$ to denote that $\pi(i) \supseteq S$ if and only if $i \in A$.

Let Π stand for the set of all choice functions. A **social aggregation function (SAF)**, is a (possibly partial) function $F : \Pi \to \mathscr{P}(T)$. $F(\pi)$ is called the **social opinion** under the SAF F. Note that since F is a function from Π to $\mathscr{P}(T)$, in this setup the set of agents (indeed Π) and the agenda are treated as being fixed for any given F. Since this never leads to much confusion, we often leave this point implicit in what follows.[11]

Axioms

A SAF transforms individual judgement sets of agents into the social opinion $S \subseteq T$. In the formulation above, the SAF is left entirely unrestricted. In practice, typically, one might want to impose certain conditions on the social aggregation procedure. For instance, it seems quite natural to argue that whenever *every* agent accepts some sentence α, α should also be also socially accepted. SAFs that satisfy this condition are called **Paretian**. Formally, a SAF is called Paretian if and only if:

$$\text{for all } \pi \text{ and all } \alpha,\ [\![\alpha]\!]_\pi = N \text{ implies } \alpha \in F(\pi).$$

Such properties of a SAF that are imposed upon it are called **axioms**. Implicitly or explicitly, axioms often reflect desires of how one expects individual judgement sets to be represented in the social opinion. Clearly, it is possible to engage in deep philosophical debates on this matter. However, it is not our intent here to provide such philosophical justifications of the axioms we introduce.[12] Three well-known—and indeed well-debated—axioms will play a special rôle in the balance of this thesis.

Definition 20. *A SAF F is said to be:*
Neutral (N) *if for all $\alpha, \beta \in T$ and all $\pi, \pi' \in \Pi$, if $[\![\alpha]\!]_\pi = [\![\beta]\!]_{\pi'}$, then $\alpha \in F(\pi) \iff \beta \in F(\pi')$.*
Monotonic (M) *if for all $\alpha \in T$ and all $\pi, \pi' \in \Pi$, whenever $[\![\alpha]\!]_\pi \subseteq [\![\alpha]\!]_{\pi'}$, then*

[11] As has been observed by many authors, a very explicit parallel with belief revision can be drawn. After all, a SAF can be viewed as an operator that merges a sequence of theories. See Eckert and Pigozzi [12] for more in depth discussion.

[12] Monotonicity has been defended on grounds of non-manipulability (E.g. Batteau *et al.* [3]). Neutrality has been defended on grounds of efficiency. For instance, Mihara [28] relates neutrality to considerations of computational complexity. Rubinstein [37] studies the relationship between neutrality and the quantifier depth of aggregation rules.

$\alpha \in F(\pi)$ *implies* $\alpha \in F(\pi')$.
*Have **Universal Domain (UD)** if the domain of* F *is* Π.

Roughly, Neutrality says that a SAF decides on each proposition in T in the same way. Monotonicity says when the number of agents in favour of some proposition α increases, this doesn't affect the social opinion on α unfavourably. Universal domain says that F always produces a decision, for every choice function π. Note in particular that the majority rule is Neutral, Monotonic, and has Universal Domain.

Besides the above three axioms, some other axioms are of interest because they make demands of a SAF F on the way it preserves the logical structure that is present in agendas. For instance, F is called **Collectively Rational** if for all π, $F(\pi)$ is consistent whenever each $\pi(i)$ is consistent. We say that a SAF F is **Cautious** if for all π and all positive α, it is never the case that $\{\alpha, \neg\alpha\} \subseteq F(\pi)$. Of course $\{\alpha, \neg\alpha\} \subseteq S$ would make the theory S inconsistent, and so this is a consistency constraint on F, but note that cautiousness demands less of F than collective rationality. A cautious SAF might still select certain other inconsistent subsets of T (For instance, the minimally inconsistent subset considered in example 19 does not contain a pair $\alpha, \neg\alpha$, and thus isn't precluded by cautiousness.) Conversely, we say that F is **Resolute** if for all π and all positive α, it is never the case $F(\pi) \cap \{\alpha, \neg\alpha\} = \varnothing$. A SAF F is called **Decisive** if for all π, $F(\pi)$ is complete, or equivalently, if F is both cautious and resolute.

Finally, we define an anonymity property. Let $\sigma : N \to N$ be a permutation of N. Denote by $\sigma\pi$ the choice function π' such that $\pi'(i) = \pi(\sigma(i))$. A SAF F is called **(Finitely) Anonymous** if for all $\pi \in \Pi$ and all $\alpha \in T$ and all (finite) permutations σ, if $\alpha \in F(\pi)$ then $\alpha \in F(\sigma\pi)$. Anonymity says that whether some $\alpha \in T$ is included in the social opinion is determined by the *number* of agents in support of it (or more precisely: by the *cardinality* of agents in support of it), and not by their identity. Finite anonymity says that the identity of any *finite* subset of agents in support of a proposition does not matter. Obviously, if N is a finite set, finite anonymity coincides with anonymity.

3.2 Linear Orderings and Condorcet Cycles

To illustrate the judgement aggregation framework and some of the material that follows, let us first consider the problem originally studied by Condorcet. This is the problem of ranking a set of alternatives through majority voting. Formally, a ranking of a set of ℓ alternatives $A = \{a_1, a_2, \ldots, a_\ell\}$ is simply a linear ordering ("$<$") of A. Recall that the theory of linear orderings is given by the following first order sentences (see [19]):

$$\forall x,\, x \not< x \tag{LO1}$$

$$\forall x \forall y,\, x = y \vee x < y \vee x > y \tag{LO2}$$

$$\forall x \forall y \forall z,\, x < y \wedge y < z \to x < z \tag{LO3}$$

Suppose that some set of agents has to decide collectively on a ranking of the alternatives in A. That is, each agent $i \in N$ ranks the alternatives according to his or her individual linear ordering $<_i$, and we are looking for a social opinion, or rather a social ranking, $<$, of A. Given L01–L03 $a_j \not< a_k$ implies $a_k < a_j$ (for $j \neq k$), and thus any linear ordering over A is completely determined by some complete subset of the following agenda:

$$T(A) := \{a_j < a_k \mid 1 \leq j < k \leq \ell\} \cup \{a_j \not< a_k \mid 1 \leq i < j \leq \ell\}.$$

We would like the social opinion on $T(A)$ to give rise to a linear order, and thus should at least require that the social opinion is consistent with respect to the theory of linear orderings LO1–LO3. It isn't hard to see that subset $S \subseteq T$ is consistent in this sense if it its model can be extended to a linear ordering, that is, if and only if the social opinion doesn't give rise to cycles (for a publication that spells out this observation in all details, see [16]). Fortunately, we can easily incorporate the idea of a "background theory" in the notion of Collective Rationality in the following way. Recall that a SAF F is Collectively Rational if for all π, $F(\pi)$ is consistent whenever for each $i \in N$, $\pi(i)$ is consistent. This notion may be relativised to a background theory $\mathscr{B}$ as follows:

Definition 21. *Let $\mathscr{B}$ be a first order theory. A SAF F is called **Collectively Rational relative to** $\mathscr{B}$, if $F(\pi) \cup \mathscr{B}$ is consistent whenever for each $i \in N$, $\pi(i) \cup \mathscr{B}$ is consistent.*

This relativised notion of Collective Rationality collapses to the ordinary one introduced earlier when we take $\mathscr{B} = \varnothing$. However, we might take $\mathscr{B}$ to be the theory of linear orders, and thus explicitly express the condition that the SAF takes individual rankings of A to a social ranking.

Now consider a voting situation with three alternatives $A' = \{a_1, a_2, a_3\}$ and a set $N' = \{1, 2, 3\}$ of three agents who decide on a collective ordering of A' by the majority rule.

$$\begin{aligned}
\pi(1) &= \{a_1 < a_2, a_1 < a_3, a_2 < a_3\} \quad &&\text{(Agent 1 orders } A' \text{ by } a_1 <_1 a_2 <_1 a_3\text{)}\\
\pi(2) &= \{a_1 \not< a_2, a_1 \not< a_3, a_2 < a_3\} \quad &&\text{(Agent 2 orders } A' \text{ by } a_2 <_2 a_3 <_2 a_1\text{)}\\
\pi(3) &= \{a_1 < a_2, a_1 \not< a_3, a_2 \not< a_3\} \quad &&\text{(Agent 3 orders } A' \text{ by } a_3 <_3 a_1 <_3 a_2\text{)}
\end{aligned}$$

It follows there is a majority for $a_1 < a_2$, a majority for $a_2 < a_3$, and a majority for $a_3 < a_1$. But the union of these statements is not compatible with the theory of linear orderings, and thus the majority rule (for three agents and three alternatives) isn't Collectively Rational relative to $\mathscr{B}$. This observation is known as *l'effet Condorcet*.

More generally, a choice function $\pi : N \to \mathscr{P}(T(A))$ is said to contain a **Condorcet cycle** if for some three distinct alternatives $a_x, a_y, a_z \in A$, there is a majority in favour of $a_x < a_y$, a majority in favour of $a_y < a_z$, and a majority in favour of $a_z < a_x$. It can be shown that, provided $|N|$ is odd, the majority rule is Collectively Rational relative to $\mathscr{B}$ on the domain:

$$\{\pi \in \Pi \mid \pi \text{ contains no Condorcet cycle}\}.$$

In what follows, we will argue that these observations provide exactly the kind of intuition to understand the lion's share of the impossibility results obtained by social choice theorists. To this effect, we will generalise the observations made by Condorcet in two ways: on the one hand we will study a larger class of SAFs, and on the other we will look at more general decision problems than those presented by the theory of linear orderings.

3.3 Aggregation Procedures Generated by Simple Games

As stated, the majority rule is Neutral, Monotonic, and has Universal Domain. In this subsection we narrow down the relation between simple games and the class of all N-M-UD-SAFs to a 1-1 correspondence. In fact, this is a straightforward generalisation of results obtained by Monjardet [29] for aggregation of tournaments,[13] now lifted into the judgement aggregation framework.

Let (N, W) be a simple game and consider the following method to construct the SAF $F_{(N,W)}$:

$$F_{(N,W)}(\pi) := \bigcup_{A \in W} \left(\bigcap_{i \in A} \pi(i) \right),$$

in words, $\alpha \in F(\pi)$ if there is some winning coalition A of (N, W) such that every agent $i \in A$ accepts α. Clearly such a SAF satisfies N, M and UD.

In this manner, every simple game gives rise to a SAF. Conversely, a SAF might admit a representation in terms of simple games. Say that F is **generated by** Ω, if $F = F_\Omega$ for some Ω. The following result shows that the N, M, UD-SAF s in fact stand in a bijective relation with the simple games.

Proposition 22. *Suppose F is a SAF that satisfies N, M, UD. Then there is a (unique) simple game Ω such that $F = F_\Omega$.*

Proof. Fix an arbitrary M-N-UD-SAF F and let N be the underlying set of agents. Call a set $A \subseteq N$ α-decisive iff $\alpha \in F(\pi)$ whenever $[\![\alpha]\!]_\pi = A$. As F is neutral, A is α-decisive if and only if there exists $\pi \in \Pi$ such that $[\![\alpha]\!]_\pi = A$ and $\alpha \in F(\pi)$. Denote by $W(\alpha)$ the family of α-decisive sets. By monotony, if B contains a α-decisive set A, then B is also and α-decisive set, so $W(\alpha)$ satisfies (M1). By neutrality, for all $\alpha, \beta \in T$, $W(\alpha) = W(\beta) =: W$. Then (N, W) is a simple game, and it is straightforward to verify F is generated by (N, W). ■

[13] Tournaments are connected antisymmetric binary relations.

Several properties of simple games pass at once to the SAF they correspond to, and vice versa.

Proposition 23. *Fix a set of agents N. Let T be an agenda and suppose $F : \Pi \to \mathscr{P}(T)$ satisfies* N, M, UD. *Let Ω be the simple game that generates F.*
(a) F is cautious iff Ω is proper.
(b) F is resolute iff Ω is strong.
(c) F is decisive iff Ω is a family of majorities.
(d) F is (finitely) anonymous iff the set of winning coalitions of Ω is closed under (finite) permutations.

Proof. Let $F = F_\Omega$ where $\Omega = (N, W)$.
(a $\Rightarrow$) Let α be a positive sentence of T. Suppose F is cautious, let $A \in W$. Clearly, there exists $\pi \in \Pi$ such that $[\![\alpha]\!]_\pi = A$ and hence $\alpha \in F(\pi)$. As individual judgement sets are complete, $[\![\neg\alpha]\!]_\pi = N - A$. As F is cautious $\neg\alpha \notin F(\pi)$. It follows $N - A \notin W$. Hence W satisfies (M2a).
(a $\Leftarrow$). Suppose $\alpha \in F_\Omega(\pi)$. Then there is $A \in W$ such that for every agent $i \in A$, $\alpha \in \pi(i)$. Then $A \subseteq [\![\alpha]\!]_\pi$. By (M1), $[\![\alpha]\!]_\pi \in W$. Let $\beta = \neg\alpha$ if α isn't of the form $\neg\gamma$ and $\beta = \gamma$ otherwise. As individual judgement sets are complete, $[\![\beta]\!]_\pi = N - [\![\alpha]\!]_\pi$. By (M2a), $N - [\![\beta]\!] \notin W$. Now suppose $\beta \in F_\Omega(\pi)$. Then there is $B \in W$ such that for every agent $i \in B$, $\beta \in \pi(i)$. Clearly $B \subseteq [\![\beta]\!]_\pi$, so by (M1) $[\![\beta]\!]_\pi \in W$, giving a contradiction.
(b $\Rightarrow$) Suppose F is resolute, and that $A \notin W$. Let α be a positive sentence of T; by neutrality there is no π such that $[\![\alpha]\!]_\pi = A$ and $\alpha \in F(\pi)$. Let $\pi \in \Pi$ be any choice function such that $[\![\alpha]\!]_\pi = A$. As individual judgement sets are complete, $[\![\neg\alpha]\!]_\pi = N - A$. As F is resolute we have $\neg\alpha \in F(\pi)$. Thus there exists $\pi \in P$ s.t. $[\![\neg\alpha]\!]_\pi = N - A$ and $\neg\alpha \in F(\pi)$. Hence $N - A$ is $\neg\alpha$-decisive, and thus $N - A \in W$. Thus W satisfies (M2b).
(b $\Leftarrow$) Suppose $\alpha \notin F_\Omega(\pi)$. Then there is no $A \in W$ such that every agent $i \in A$ accepts α. In particular, $[\![\alpha]\!]_\pi \notin W$. Let $\beta = \neg\alpha$ if α isn't of the form $\neg\gamma$ and $\beta = \gamma$ otherwise. As individual judgement sets are complete, $[\![\beta]\!]_\pi = N - [\![\alpha]\!]_\pi$. By (M2b), $[\![\beta]\!]_\pi \in W$. Whence $\beta \in F_\Omega(\pi)$.
(c) follows from (a) and (b).
(d $\Rightarrow$). Let $A \subseteq N$ and $\alpha \in T$ and σ a (finite) permutation. Let π be such that $[\![\alpha]\!] = A$. $A \in W$ if and only if $\alpha \in F(\pi)$ if and only if $\alpha \in F(\sigma\pi)$ if and only if $[\![\alpha]\!]_{\sigma\pi} \in W$ if and only if $\sigma[A] \in W$.
(d $\Leftarrow$). Let $\alpha \in T$, $\pi \in \Pi$, and σ a (finite) permutation. $\alpha \in F(\pi)$ if and only if $[\![\alpha]\!]_\pi \in W$ if and only if $\sigma[[\![\alpha]\!]]_\pi \in W$ if and only if $[\![\alpha]\!]_{\sigma\pi} = \sigma[A] \in W$ if and only if $\alpha \in F(\sigma\pi)$. ■

Let us briefly revisit the Simple Majority Games, and their generalisations in the form of majority spaces. A SAF F is called **Positively Responsive** if for all $\pi \in \Pi$

and $\pi' \in \Pi$, and all pairs $(\alpha, \neg\alpha)$ of the agenda T, whenever $\alpha \in F(\pi)$ and $[\![\alpha]\!]_\pi \subset [\![\alpha]\!]_{\pi'}$, then $\neg\alpha \notin F(\pi')$ and whenever $\neg\alpha \in F(\pi)$ and $[\![\neg\alpha]\!]_\pi \subset [\![\alpha]\!]_{\pi'}$, then $\alpha \notin F(\pi')$. Thus a positively responsive F is "almost" cautious: if π is such that $\{\alpha, \neg\alpha\} \in F(\pi)$, then a slight change of the set of agents that accept α or $\neg\alpha$ induces F to remove either α or $\neg\alpha$ from the set of socially accepted sentences.

A celebrated theorem of May [27] characterises the majority rule as the unique SAF that is Neutral, Resolute, Anonymous and Positively Responsive and has Universal Domain.[14] We can state the following variant of May's theorem (drawing on some results from Pacuit and Salame [33]), that generalises this characterisation when the set of agents N is infinite:

Proposition 24. *Let F be a SAF satisfying UD. The following are equivalent:*
(a) F is neutral, resolute, finitely anonymous, and positively responsive;
(b) F is neutral, resolute, monotonic, finitely anonymous, and positively responsive;
(c) F is generated by a simple game Ω and Ω is a majority space.

This yields May's theorem as a consequence, since the only majority space over a finite set N is the Simple Majority Game.

Proof. (a $\Rightarrow$ b). Suppose $\alpha \in F(\pi)$; without loss of generality we assume that α is positive sentence. Let π' be such that $[\![\alpha]\!]_\pi \subseteq [\![\alpha]\!]_{\pi'}$. If $[\![\alpha]\!]_\pi = [\![\alpha]\!]_{\pi'}$ then by neutrality $\alpha \in F(\pi)$. If $[\![\alpha]\!]_\pi \subset [\![\alpha]\!]_{\pi'}$ then, since F is positively responsive, $\neg\alpha \notin F(\pi)$. Since F is resolute, $\alpha \in F(\pi)$. So F is monotonic.
(b $\Rightarrow$ c). F is neutral, resolute, finitely anonymous, and hence by the previous lemma, F is generated by a strong simple game closed under finite permutations. It suffices to prove that positive responsiveness implies (M4).

Let (N, W) be the simple game that generates F. Suppose there are $A, B \in W$ such that $A \cap B = \varnothing$; and suppose $A \neq N - B$. We will derive a contradiction. Under our assumption there is $x \in N - (A \cup B)$ and $A' = A \cup \{x\} \in W$ by (M1). Let $\alpha \in T$ be a positive sentence. Let π be such that $[\![\alpha]\!]_\pi = A$ and hence $[\![\neg\alpha]\!]_\pi = N - A$. A is a majority, and $N - A$ contains B so is a majority, it follows that $\{\alpha, \neg\alpha\} \subseteq F(\pi)$. Let π' be such that $[\![\alpha]\!]_{\pi'} = A'$ and hence $[\![\neg\alpha]\!]_{\pi'} = N - A'$. Then A' is a majority, and $N - A'$ contains B so is a majority, it follows that $\neg\alpha \in F(\pi')$, contradicting positive responsiveness.
(c $\Rightarrow$ a). Obvious. ∎

3.4 Duality Theory for Aggregation Procedures

The literature on judgement aggregation has traditionally focused on selecting *complete* subsets of the agenda T.[15] In this subsection, we will therefore analyse SAFs $F : \Pi \to$

[14] When there are only two alternatives, Neutrality corresponds to self-duality.
[15] There are some notable exceptions in the literature. In Gärdenfors [14], Dietrich and List [9] and Daniëls and Pacuit [6], the requirement that social opinions should be complete is relaxed.

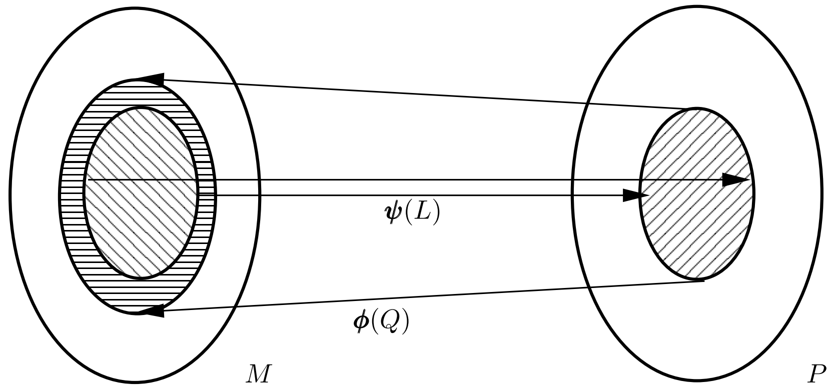


Figure 2: The Galois Connection

$\mathscr{P}(T)$ that determine such complete subsets of T, that is, Decisive SAFs. Moreover, we impose Neutrality, Monotonicity, and Universal Domain. As we have seen, such SAFs correspond to the families of majorities over N, the set of which we shall denote by M.

Let $\mathscr{B}$ be a consistent first order background theory, which we will keep fixed for the purpose of this subsection. Recall that a SAF F is called Collectively Rational relative to $\mathscr{B}$ if for all choice functions π, $F(\pi) \cup \mathscr{B}$ is consistent whenever each $\pi(i) \cup \mathscr{B}$ is consistent. In addition to $\mathscr{B}$, fix an agenda T and a set of agents N, and let us denote the set of all choice functions such that $\pi(i) \cup \mathscr{B}$ is consistent for all $i \in N$ by $P \subseteq \Pi$.

Let us call a pair $\pi \in P$ and $\Omega \in M$ **admissible** if $F_\Omega(\pi) \cup \mathscr{B}$ is consistent. We can use this notion to establish a duality relation between the lattice of sets profiles $(\mathscr{P}(P), \subseteq)$ on the one hand, and the lattice of sets of families of majorities $(\mathscr{P}(M), \subseteq)$, on the other. We may define the following operation on $\mathscr{P}(P)$:

$$\boldsymbol{\phi}(Q) := \{\Omega \in M \mid \text{for all } \pi \in Q,\ (\pi, \Omega) \text{ is admissible}\}.$$

$\boldsymbol{\phi}$ takes a set $Q \subseteq P$ to the largest set $L \subseteq M$ of families of majorities such that each $F_\Omega(\pi)$, where $\Omega \in L$ and $\pi \in Q$, is Collectively Rational relative to $\mathscr{B}$. Dually, we define the operation $\boldsymbol{\psi}$ on $\mathscr{P}(M)$:

$$\boldsymbol{\psi}(L) := \{\pi \in P \mid \text{for all } \Omega \in L,\ (\pi, \Omega) \text{ is admissible}\}.$$

$\boldsymbol{\psi}$ takes a set $L \subseteq M$ to the largest domain of choice functions such that each F generated from L is Collectively Rational relative to $\mathscr{B}$ on the entire domain. Figure 2 depicts these operations on M and P. The operations $\boldsymbol{\psi}, \boldsymbol{\phi}$ set up a Galois connection between the lattices $(\mathscr{P}(P), \subseteq)$ and $(\mathscr{P}(M), \subseteq)$. This fact hinges on no particular feature of our definition of admissibility—it is simply a more general result on a binary relation between two sets, in the present case M and P and the admissibility-relation. For reference, we record some well known properties of Galois connections between ordered sets (See Davey and Priestley [7] for discussion, in particular lemma 7.26):

$$Q \supseteq Q' \implies \boldsymbol{\phi}(Q) \supseteq \boldsymbol{\phi}(Q') \qquad Q \supseteq \boldsymbol{\psi} \circ \boldsymbol{\phi}(Q).$$
$$L \supseteq L' \implies \boldsymbol{\psi}(L) \supseteq \boldsymbol{\psi}(L') \qquad L \subseteq \boldsymbol{\phi} \circ \boldsymbol{\psi}(L).$$

$$\boldsymbol{\psi} \circ \boldsymbol{\phi} \text{ and } \boldsymbol{\phi} \circ \boldsymbol{\psi} \text{ are closure operators}$$

Our next result shows that the problem of aggregation becomes of interest if the agenda itself is "interesting enough"; in the sense that the propositions under collective scrutiny are logically intertwined in the following, precise, sense. Let us call an agenda **linked** if it contains a set X^*, where $|X^*| \geq 3$, such that $X^* \cup \mathscr{B}$ is inconsistent, and for each proper subset $Y \subset X^*$, $Y \cup \mathscr{B}$ is consistent. That is, X^* is minimally consistent given the background theory $\mathscr{B}$, and contains at least three sentences.

Lemma 25. *The following are equivalent:*
(1) $\boldsymbol{\psi}(M) \neq P$*;*
(2) $\boldsymbol{\phi}(P) \neq M$*;*
(3) The agenda T *is linked.*

Proof. We argue by contraposition.
$(2 \Rightarrow 1)$. Assume $\boldsymbol{\psi}(M) = P$. Clearly $M \supseteq \boldsymbol{\phi}(P)$. Since $\boldsymbol{\phi}, \boldsymbol{\psi}$ form a Galois connection, $M \subseteq \boldsymbol{\phi}(\boldsymbol{\psi}(M))$. $\boldsymbol{\psi}(M) \subseteq P$ implies $\boldsymbol{\phi}(\boldsymbol{\psi}(M)) \subseteq \boldsymbol{\phi}(P)$. Hence $M \subseteq \boldsymbol{\phi}(\boldsymbol{\psi}(M)) \subseteq \boldsymbol{\phi}(P) \subseteq M$, as required.
$(3 \Rightarrow 2)$. Suppose $\boldsymbol{\phi}(P) = M$. Then there is no $\pi \in P$ and $\Omega \in M$ such that $F_\Omega(\pi) \cup \mathscr{B}$ is inconsistent. Suppose that T contains a subset X^* that is minimal inconsistent given $\mathscr{B}$, such that $|X^*| \geq 3$; we will derive a contradiction. Consider the dualton game of example 8. Let α_i, α_j, α_k be two different sentences of X^*. Since X^* is minimally inconsistent given $\mathscr{B}$, by assumption each proper subset of X^* is consistent with $\mathscr{B}$ and thus has a model. Accordingly, pick $\pi \in P$ such that $\pi(i) \supset X^* - \{\alpha_i\}$, $\pi(j) \supset X^* - \{\alpha_j\}$ and $\pi(k) \supset X^* - \{\alpha_k\}$. Then for each $\alpha \in X^*$, $[\![\alpha]\!]_\pi \supseteq A$ for some $A \in \{\{i,j\}, \{i,k\}, \{j,k\}\}$, and hence $\alpha \in F(\pi)$. This means $X^* \subseteq F(\pi)$, a contradiction.
$(1 \Rightarrow 3)$. Assume T doesn't contain a set X^* such that $|X^*| \geq 3$ and that is minimally inconsistent given $\mathscr{B}$. Suppose some for some $\pi \in P$ and $(N, W) \in M$, $(\pi, (N, W))$ isn't consistent. Then $F_{(N,W)}(\pi)$ contains a set S^*, such that $S^* \cup \mathscr{B}$ is inconsistent. By compactness there is a finite such S^*, and hence there is such a set S^* if smallest cardinality. Clearly, in this case, $|S^*| < 3$. Moreover, $|S^*| > 1$. If not, $S^* = \{\alpha\}$, $\alpha \in F_{(N,W)}(\pi)$, and hence $[\![\alpha]\!]_\pi \in W$, implying $[\![\alpha]\!]_\pi \neq \varnothing$, and thus $\alpha \in \pi(i)$ for some $i \in [\![\alpha]\!]_\pi$. But then $\pi(i) \cup \mathscr{B}$ is inconsistent, contradicting $\pi \in P$. Therefore, $|S^*| = 2$ and thus $S = \{\alpha, \beta\}$. Hence $[\![\alpha]\!]_\pi \in W$. As individual judgement sets are consistent with $\mathscr{B}$ and $\{\alpha, \beta\}$ isn't, it must be that $[\![\beta]\!]_\pi \subseteq N - [\![\alpha]\!]_\pi$. But then $[\![\beta]\!]_\pi \cap [\![\alpha]\!]_\pi = \varnothing$, and hence by (M3), $[\![\beta]\!]_\pi \notin W$, thus $\beta \notin F_{(N,W)}(\pi)$, again a contradiction. ∎

Note that if the agenda T contains a set X^* such that $X^* \cup \mathscr{B}$ is inconsistent, then it contains a finite set X^* that is minimally inconsistent (given the background theory $\mathscr{B}$), because first order logic is compact.

Generalised Condorcet Cycles

Let $\pi : N \to \mathscr{P}(T)$ be a choice function. Of course, given the choice function, π, any given SAF F determines a subset S of T. On the other hand, given π, a set $S \subseteq T$ determines a family of subsets of N in the following way:

$$\mathscr{A}(\pi, S) := \{[\![\alpha]\!]_\pi \in \mathscr{P}(N) \mid \alpha \in S\}.$$

Observe that if (N, W) is any simple game such that $\mathscr{A}(\pi, S) \subseteq W$, then $S \subseteq F_{(N,W)}(\pi)$. Does such a simple game exist? The answer is yes, because we can obtain a simple game simply by taking the closure of $\mathscr{A}(\pi, S)$ under (M1). Thus (π, S) determines a non-empty collection of monotonic, neutral, universal domain SAFs, $\mathscr{F}(\pi, S)$.

Given our focus on families of majorities, we seek properties of $\mathscr{A}(\pi, S)$ such that $\mathscr{F}(\pi, S)$ contains a family of majorities. Now lemmas 12 and 13 immediately tell us that if $\mathscr{A}(\pi, S)$ has the PIP, then $\mathscr{F}(\pi, S)$ contains families of majorities and if $\mathscr{A}(\pi, S)$ has the FIP, then $\mathscr{F}(\pi, S)$ contains ultrafilters. These observations motivate the following definition.

Definition 26. *Fix an agenda T. We say that a choice function $\pi \in P$ contains a* ***generalised Condorcet cycle*** *if there is a set $S^* \subseteq T$ such that:*
(a) $|S^*| \geq 3$;
(b) $\mathscr{A}(\pi, S^*)$ has the FIP;
(c) $S^* \cup \mathscr{B}$ is inconsistent, and for each proper subset $S \subset S^*$, $S \cup \mathscr{B}$ is consistent.

The analogy with the Condorcet cycles introduced earlier should be clear. Let P^* denote the set of all π that contain no generalised Condorcet cycles and M^* denote the set of all ultrafilters.

A set $Q \subseteq P$ is called closed if $Q = \boldsymbol{\psi}(\boldsymbol{\phi}(Q))$, and similarly a set $L \subseteq M$ is closed if $L = \boldsymbol{\phi}(\boldsymbol{\psi}(L))$. Denote by $P^* \subseteq P$ the set of all profiles that contain no generalised condorcet cycle of length 3 or greater; By $M^* \subseteq M$ the ultrafilters. The following proposition characterises some important closed subsets of P and M. In fact, from this result we will deduce most of the impossibility results stated in the next subsection.

Proposition 27. *Let T be linked.*

(a) $\boldsymbol{\phi}(P) = \boldsymbol{\phi}(P - P^*) = M^*$	*(b)* $\boldsymbol{\psi}(M^*) = P$
(c) $\boldsymbol{\psi}(M) = \boldsymbol{\psi}(M - M^*) = P^*$	*(d)* $\boldsymbol{\phi}(P^*) = M$

Proof. Claim 1. ($P^* \subseteq \boldsymbol{\psi}(M)$). Let $\Omega \in M$ and $\pi \in P$ be such that $F_\Omega(\pi) \cup \mathscr{B}$ is inconsistent. We will show that $\pi \notin P^*$. Let $\Omega = (N, W)$.

As $F_\Omega(\pi) \cup \mathscr{B}$ is inconsistent, (by compactness) $F_\Omega(\pi)$ contains a finite subset S^* that is minimally inconsistent given $\mathscr{B}$.

We claim $|S^*| > 1$. If not, $S^* = \{\alpha\}$ for some $\alpha \in T$, and hence $[\![\alpha]\!]_\pi \in W$, implying $[\![\alpha]\!]_\pi \neq \varnothing$, and thus $\alpha \in \pi(i)$ for some $i \in [\![\alpha]\!]_\pi$, which means $\pi(i) \cup \mathscr{B}$ is inconsistent, contradicting $\pi \in P$. Thus $|S^*| \geq 2$.

We claim $|S^*| > 2$. Suppose not. Then $|S^*| = 2$ and thus $S^* = \{\alpha, \beta\}$ for some $\alpha, \beta \in T$. It must be that $[\![\alpha]\!]_\pi \in W$. As individual judgement sets are consistent with $\mathscr{B}$ and $\{\alpha, \beta\}$ isn't, it must also be the case that $[\![\beta]\!]_\pi \subseteq N - [\![\alpha]\!]_\pi$. But then $[\![\beta]\!]_\pi \cap [\![\alpha]\!]_\pi = \varnothing$, and hence by (M3), $[\![\beta]\!]_\pi \notin W$, and thus $\beta \notin F_{(N,W)}(\pi)$, again a contradiction.

As $S^* \subseteq F_\Omega(\pi) \subseteq T$, for each $\alpha \in S^*$, $[\![\alpha]\!]_\pi \in W$. Therefore, since Ω is a family of majorities, the set $\{[\![\alpha]\!]_\pi \mid \alpha \in S^*\}$ has the PIP. This proves Claim 1.

Claim 2. ($\boldsymbol{\psi}(M^*) = P$). Let $(N, W) \in M^*$ and $\pi \in P$ and suppose $F_{(N,W)}(\pi) \cup \mathscr{B}$ is inconsistent; we will derive a contradiction. Then (by compactness) there is a finite subset $S^* \subseteq F_{(N,W)}(\pi)$ such that S^* is minimally inconsistent given $\mathscr{B}$. Now:

$$\begin{aligned}
& \forall \alpha \in S^*, \alpha \in F_{(N,W)}(\pi) \\
\iff\quad & \forall \alpha \in S^*, [\![\alpha]\!]_\pi \in W \\
\implies\quad & \bigcap_{\alpha \in S^*} [\![\alpha]\!]_\pi \neq \varnothing \quad \text{(W has the } \mathsf{FIP}) \\
\iff\quad & \exists i \in \bigcap_{\alpha \in S^*} [\![\alpha]\!]_\pi \\
\iff\quad & S^* \subseteq \pi(i).
\end{aligned}$$

But this is impossible: since $\pi \in P$, $\pi(i) \cup \mathscr{B}$ is consistent.

Claim 3. ($\boldsymbol{\psi}(M - M^*) \subseteq P^*$). We will show that for any $\pi \notin P^*$, there exists $\Omega \in M - M^*$ such that (π, Ω) isn't admissible. Pick such $\pi \notin P^*$; then π contains a generalised condorcet cycle over some set S^*. Let $W = \mathscr{A}(\pi, S^*)$. By Lemma 13, W can be extended to a family of majorities $\Omega = (N, W^*)$. We wish to argue that Ω isn't an ultrafilter. Suppose it is; then W^* has the FIP and therefore $\mathscr{A}(\pi, S)$ has the FIP. Since (by compactness) S^* is a finite set there is $i \in \bigcap \mathscr{A}(\pi, S^*)$. It follows that $S^* \subseteq \pi(i)$, which is absurd. Hence W^* doesn't have the FIP, and isn't an ultrafilter. Thus $\Omega \in M - M*$. Clearly $S^* \subseteq F_\Omega(\pi)$ and thus (π, Ω) isn't admissible, proving the claim.

Claim 4. ($\boldsymbol{\phi}(P - P^*) \subseteq M^*$). We will show that for any $\Omega \notin M^*$, there exists $\pi \in P - P^*$ such that (π, Ω) isn't admissible. By Lemma 14, $(N, W) \notin M^*$ if and only if there exists a triple $A_1, A_2, A_3 \in W$ such that $A_1 \cap A_2 \cap A_3$ is empty. T is linked, and thus contains a set $S^* \subseteq T$, such that S^* is minimally consistent given $\mathscr{B}$. Let α_1, α_2 be two different sentences of S^*. Since S^* is minimally inconsistent, by assumption each proper subset is consistent and has a model. Accordingly, let $\pi \in P$ such that for each $i \in A_1$, $\pi(i) \supset \{\alpha_1\}$, for each $i \in A_2$, $\pi(i) \supset \{\alpha_2\}$ and for each $i \in A_3$, $\pi(i) \supset X^* - \{\alpha_1, \alpha_2\}$. Then for each $\alpha \in S^*$, $[\![\alpha]\!]_\pi \supseteq A_j$ for some $j \in \{1, 2, 3\}$, and hence $\alpha \in F_{(N,W)}(\pi)$, and thus $S^* \subseteq F_{(N,W)}(\pi)$. Hence $(\pi, (N, W))$ isn't admissible, as claimed.

(a). Clearly $P - P^* \subseteq P$. As $\boldsymbol{\phi}$ is antitonic, it follows $\boldsymbol{\phi}(P) \subseteq \boldsymbol{\phi}(P - P^*)$. Moreover, by claim 2, $P \subseteq \boldsymbol{\psi}(M^*)$. Since $\boldsymbol{\psi}, \boldsymbol{\phi}$ is a Galois connection, $\boldsymbol{\phi}(P) \supseteq M^*$. Finally, by claim 4 $\boldsymbol{\phi}(P - P^*) \subseteq M^*$. Thus $M^* \subseteq \boldsymbol{\phi}(P) \subseteq \boldsymbol{\phi}(P - P^*) \subseteq M^*$.

(b). This is claim 2.

(c). $M - M^* \subseteq M$. So $\boldsymbol{\psi}(M) \subseteq \boldsymbol{\psi}(M - M^*)$. By claims 1 and 3 we find $P^* \subseteq \boldsymbol{\psi}(M) \subseteq \boldsymbol{\psi}(M - M^*) \subseteq P^*$.

(d). Clearly $\boldsymbol{\phi}(P^*) \subseteq M$. By claim 1, $P^* \subseteq \boldsymbol{\psi}(M)$. Since $\boldsymbol{\psi}, \boldsymbol{\phi}$ is a Galois connection, $\boldsymbol{\phi}(P^*) \supseteq M$. ■

3.5 Arrovian Theorems

In essence, proposition 27 establishes a relation between SAFs and the domain on which they are consistent. Since $\boldsymbol{\phi}(P^*) = M$ and $P^* = \boldsymbol{\psi}(M)$, an arbitrary monotonic, neutral, and decisive aggregation procedure is collectively rational over all domains of choice functions that do not contain generalised condorcet cycles. On the other hand, imposing universal domain, we find that a monotonic, neutral, and decisive aggregation procedure gives a consistent outcome for every choice function if and only if it is generated by an ultrafilter; this observation follows from $\boldsymbol{\psi}(M^*) = P$ and $M^* = \boldsymbol{\phi}(P)$.

Using the left-to-right direction of the last observation, we have established the following. A SAF is called **Dictatorial** if there is $i \in N$ such that for all π, $F(\pi) = \pi(i)$.

Theorem 28. *Let $\mathscr{B}$ be a consistent first order theory. Let N be finite, T be linked, and $F : \Pi \to \mathscr{P}(T)$ be a Decisive SAF that satisfies Neutrality, Monotonicity, and Collective Rationality relative to $\mathscr{B}$, and has universal domain. Then F is Dictatorial.*

Proof. By proposition 27 and the fact that all ultrafilters over finite sets N are dictatorships. ■

Moreover, the right-to-left direction of the observation should not come as a surprise to logicians, since it is an easy consequence of the well known theorem of Łoś on ultraproducts (see [19]).

The duality result established in the previous subsection holds for all linked agendas. Clearly then, impossibility results such as theorem 28 emerge from an interplay between constraints induced axioms on the one hand, and by the logical structure of the agenda on the other. The structure of the agenda can be more intricate than demanded by the property of linked-ness. We will investigate some consequences of the structure of the agenda properties (taken from Dietrich and List, [10]):

Definition 29. *Let $\mathscr{B}$ be a consistent first order theory. Define $\neg Y := \{\neg\alpha \mid \alpha \in Y\}$.*
*(a) An agenda T is called **symmetric** if for all $Y \subset T$, whenever $Y \cup \mathscr{B}$ is consistent, then so is $\neg Y \cup \mathscr{B}$.*
*(b) An agenda T is called **balanced** if there is a set $X^{**} \subseteq T$ such that $X^{**} \cup \mathscr{B}$ is inconsistent, and a tripartition A^*,B^*,C^* of X^{**} such that $A^* \cup \neg B^* \cup \neg C^* \cup \mathscr{B}$ is consistent, $A^* \cup \neg B^* \cup C^* \cup \mathscr{B}$ is consistent, and $A^* \cup B^* \cup \neg C^* \cup \mathscr{B}$ is consistent.*

In Dietrich and List [10], an agenda is called minimally connected if it is both linked and balanced. To define this notion of minimally connectedness, they state a property that is (in our view) slightly more baroque than our notion of balancedness, but which in fact implies it. Unsurprisingly, then, the following result mirrors parts of Theorem 1 and Proposition 1 in Dietrich and List [10]:

Lemma 30. *Let $\mathscr{B}$ be a consistent first order theory. Suppose F is a Resolute SAF that satisfies Neutrality, Collective Rationality relative to $\mathscr{B}$, and has Universal Domain.*
(a) If the agenda is asymmetric then F is necessarily Paretian;
(b) If F is Paretian and the agenda is linked and balanced, then F is necessarily Monotonic (thus generated by a strong simple game);
(c) If the agenda is linked, balanced and asymmetric then F is Paretian and Monotonic (thus generated by a strong simple game).

Proof. (a) Suppose T is asymmetric, and let $Y \subseteq T$ be a set such that $Y \cup \mathscr{B}$ is consistent but $\neg Y \cup \mathscr{B}$ isn't. Let $\pi \in P$ be a choice function such that for all $i \in N$, $\pi(i) \supseteq Y$. Pick $\alpha \in Y$. F is Neutral, so if $\alpha \in F(\pi)$, then F is Paretian. On the other hand, if $\alpha \notin F(\pi)$, then $\neg\alpha \in F(\pi)$, as F is Resolute. By Neutrality, $\neg Y \subseteq F(\pi)$. This means $F(\pi) \cup \mathscr{B}$ is inconsistent, contradicting Collective Rationality relative to $\mathscr{B}$.
(b) Fix an arbitrary M-N-UD-SAF F and let N be the underlying set of agents. Call a set $A \subseteq N$ α-decisive iff $\alpha \in F(\pi)$ whenever $[\![\alpha]\!]_\pi = A$. As F is Neutral, A is α-decisive if and only if there exists $\pi \in \Pi$ such that $[\![\alpha]\!]_\pi = A$ and $\alpha \in F(\pi)$. Denote by $W(\alpha)$ the family of α-decisive sets. By Neutrality, for all $\alpha, \beta \in T$, $W(\alpha) = W(\beta) =: W$.

We will show that W is upwards closed (that is, satisfies (M1)). Let $X^{**} \subseteq T$ be a set as stated in definition 29. Suppose $A \subseteq B$ and $A \in W$. Consider π such that:

$[\![A^* \cup C^*]\!]_\pi = A$, $[\![A^* \cup \neg C^*]\!]_\pi = N - A$, $[\![A^* \cup \neg B^*]\!]_\pi = B$, and $[\![A^* \cup B^*]\!] = N - B$.

Note that there exists such a π such that $\pi(i) \cup \mathscr{B}$ is consistent for all $i \in N$ (in particular $i \in (N - A \cap B)$ implies $\pi(i) \supseteq A^* \cup \cup \neg B^* \neg C^*$ which is consistent with $\mathscr{B}$ by assumption).

Since $A \in W$, $C^* \subseteq F(\pi)$. Since F is Paretian, $A^* \subseteq F(\pi)$. Suppose $B \in W$. Then $A^* \cup B^* \cup C^* \subseteq F(\pi)$, contradicting Collective Rationality relative to $\mathscr{B}$. F is Resolute and Neutral, so $\neg B^* \subseteq F(\pi)$. Therefore $B \in W$ and so W is closed under supersets. By proposition 24 F is generated by a strong simple game.
(c) follows from (a) together with (b). ∎

A SAF is called **Persecutive** if there is $i \in N$ such that for all π, $F(\pi) = T - \pi(i)$. Combining the results of this and the previous subsection, we have Dietrich and List's counterpart to Arrow's famous impossibility theorem [10]:

Theorem 31. *Let $\mathscr{B}$ be any consistent first order theory. Let N be finite. Let the agenda be linked and balanced. Let F be a SAF satisfying UD. The following are equivalent:*
(a) F is Collectively Rational relative to $\mathscr{B}$, Decisive, and Neutral;
(b) F is Dictatorial, or F is Persecutive and the agenda is symmetric.

Proof. We prove the non-obvious direction $(a) \implies (b)$. If F is Paretian, we find that F is Monotonic by lemma 30, and whence F is generated by a family of majorities Ω, and must be Ω an ultrafilter, by preceding results. If N is finite, F is Dictatorial.

If F isn't Paretian, we find that the agenda is symmetric by lemma 30. In this case, define $F^\partial = X - F(\overline{\pi})$. Clearly F^∂ is Paretian, Decisive, Neutral and Consistent. Whence by lemma 30, F^∂ is Monotonic. Thus F^∂ is generated by a family of majorities Ω, and Ω must be an ultrafilter, by preceding results. If N is finite, F^∂ is dictatorial, and thus $F = X - F^\partial(i) = X - \pi(i)$ is persecutive. ∎

The Majority Rule

To conclude this section, let us look once more at the relationship between the general judgement aggregation framework and its application to aggregating linear orderings. It is worth noting that for aggregating linear orderings over a set A of alternatives, every agenda is linked, balanced, and symmetric provided that $|A| > 3$. After all, in this case the agenda contains the minimally inconsistent set $X^* = \{a_1 < a_2, a_2 < a_3, a_3 < a_1\}$.[16] Clearly, $\{a_1 < a_2, a_2 < a_3, a_1 < a_3\}$ is consistent, as are $\{a_1 < a_2, a_3 < a_2, a_1 < a_3\}$ and $\{a_1 < a_2, a_3 < a_2, a_3 < a_1\}$, so T is balanced. Moreover, if a set $Y \subseteq T$ contains no cycles, then the set $\neg Y = \{a < b \mid b < a \in Y\}$ contains no cycles.

Eric Maskin proved a remarkable result on the robustness of the majority rule [26] in this classical Arrovian context. Maskin's theorem says that, for odd $|N|$, and for

[16] To make our notation more transparent, we simply write $a_3 < a_1$ instead of $a_1 \not< a_3$; given the background theory of linear orderings, these statements are equivalent.

aggregating linear orderings, the majority rule yields linear orderings as social opinions at least as often as any other social aggregation procedure that is Neutral, Resolute, Paretian, Anonymous, and has Universal Domain.

Maskin's theorem readily lifts into the judgement aggregation framework. Let us say that given a class K of $\mathsf{UD\text{-}SAF}$s, a SAF $F \in \mathsf{K}$ is **maximally consistent for** K if for all $F' \in \mathsf{K}$, and for all π, if $F'(\pi)$ is consistent, then $F(\pi)$ is consistent. Similarly, for any first order theory $\mathscr{B}$, $F \in \mathsf{K}$ is **maximally consistent for** K **relative to** $\mathscr{B}$ if for all $F' \in \mathsf{K}$, and for all π, if $F'(\pi) \cup \mathscr{B}$ is consistent, then $F(\pi) \cup \mathscr{B}$ is consistent.

Theorem 32. *Let T be any agenda, $\mathscr{B}$ be any consistent first order theory, let N be finite and let K be the class of all Resolute, Monotonic, Neutral, and Anonymous $\mathsf{UD\text{-}SAF}$s over N. The following are equivalent:*
(a) F is maximally consistent for K relative to $\mathscr{B}$;
(b) F is the majority rule.

Hence, no Resolute, Monotonic, Anonymous, and Neutral SAF improves on the majority rule. If T is balanced, then the monotonicity condition in the theorem can be replaced by the Pareto property, by lemma 30. Actually, by resorting to simple games, the proof below is quite straightforward and takes away much of the mystery associated with Maskin's original theorem.

Proof. (a $\Rightarrow$ b). Let F' be the majority rule. The majority rule is Resolute, Monotonic, Neutral, and Anonymous, and so $F' \in \mathsf{K}$. Let (N, W') be the simple majority game over N that generates the simple majority rule. Let F be any SAF that satisfies the conditions of the theorem and is maximally consistent. By proposition 22, F is generated by a simple game (N, W); by proposition 23, W is closed under finite permutations. Suppose (N, W) doesn't satisfy (M4) and hence there are $A, B \in W$ such that $B \cap A = \varnothing$ and that $B \neq N - A$; we will derive a contradiction. Since (N, W') satisfies (M2b), either $A \in W$ or $B \in W$. Moreover, since (N, W') satisfies (M4), $A \in W$ implies $B \notin W$ and vice versa. Without loss of generality we may assume $A \in W$. Pick a positive formula $\alpha \in T$. Let $A^* \subseteq T$ be a consistent (given $\mathscr{B}$) and complete set containing α and let B^* be a consistent (given $\mathscr{B}$) and complete subset of T containing $\neg\alpha$. Let $[\![A^*]\!]_\pi = N - B$ and $[\![B^*]\!] = B$. Since $A \subseteq N - B$, $A \in W'$ implies $N - B \in W'$ by Monotony. By Neutrality and the fact that $B \notin W'$, $F'(\pi) = A^*$ and so $F'(\pi) \cup \mathscr{B}$ is consistent. By Neutrality and the fact that $B \in W$, $F(\pi) = A^* \cup B^*$. But this implies $\{\alpha, \neg\alpha\} \subseteq F(\pi)$, and so $F(\pi)$ is inconsistent. So F isn't maximally consistent, a contradiction. Conclude that F satisfies (M4) after all; by proposition 24 F is generated by a majority space, and since N is finite this means F is just the majority rule.

(b $\Rightarrow$ a). Let $F' \in \mathsf{K}$. By proposition 22, F' is generated by a strong simple game (N, W'), and W' is closed under finite permutations by proposition 23. Let (N, W)

be the simple majority game for N, generating the majority rule F. By proposition 9 $W \subseteq W'$. Whence $F_{(N,W)}(\pi) \subseteq F'_{N,W'}$. Now, the claim follows easily by contraposition: suppose $F_{N,W}(\pi) \cup \mathscr{B}$ is inconsistent, then $F'(\pi) \cup \mathscr{B}$ must be inconsistent as well. Since F' was arbitrary, it follows F is maximally consistent. ■

4 The Logic of Simple Games

The results of the previous section show that logical consistency at the level of social opinions is typically too much to demand, even if individual judgements are consistent. Still, social opinions obey certain logical rules. In this part we investigate what such rules look like. To this end, we will consider a formal logical language which can express things about social opinions. This language should take into account two facets of judgement aggregation. First, the rules that govern social opinions will typically depend on the particular aggregation procedure that is used to arrive at the social opinion. Second, as we have seen, judgement aggregation might produce inconsistent social opinions. Pauly [34],[35] has introduced a minimalistic modal-flavoured logic of collective decision making that is able to deal with these two issues. The properties of this logic will be our object of study in this section.

In contrast with the previous section, we will not be concerned with impossibility results *per se*, but rather studying the methodology of social choice theorists using methods familiar to logicians. Why might such an investigation be of interest to logicians and/or social choice theorists? From the logical stance the answer is simply that logics of aggregation procedures are interesting objects of study. From the perspective of social choice theory, logical formulae expressing properties of social aggregation procedures behave just like axioms. A logic of collective agency gives us a precise framework to investigate the rôle of axioms.

When logicians think of axiomatising a class of structures by a class of sentences, they usually have a specific language in mind, perhaps first order logic. Social choice theorists usually speak of axiomatisations when they provide a collection of properties that characterises a class of aggregation procedures, and this approach is independent of any language specific constraints. To see how these two approaches to axiomatisations tie up, much of our analysis in this section is focused on the expressive strength of the language of collective decision making. In other words, we investigate what kind of properties of social aggregation procedures can be represented in our language by viewing SAFs as mathematical structures.

The focus on a particular logical language leads us to entertain slightly different questions than are usually found in the literature social choice theory. We show that investigations of the expressive power of the language can indeed provide insights about the social choice theorist's axiomatic method. For instance, we show that certain kinds of axioms (namely the ones that can be represented in the language) cannot express non-dictatorship; in effect a very simple kind of "impossibility theorem". Furthermore, to effectively delineate the expressive power of the language, we work towards simple characterisation results that show how the minimalistic logic fits into the larger picture of modal logic, which is a strictly stronger language.

4.1 Preliminaries

Pauly's majority logic is a minimalistic language that is able to express the "social acceptance" of certain sets of propositions. For the purpose of our investigation, these propositions will be sentences uttered in a basic language $\mathscr{L}_c$, that we will simply take to be classical propositional logic. Thus, formulae in the language $\mathscr{L}_c$ are constructed from a set of sentence letters $q_1, q_2, \ldots$, and the logical connectives $\wedge, \neg$. Throughout the section we follow the standard conventions for bracketing and use the abbreviations $\rightarrow$, $\leftrightarrow$, $\vee$. By $\models$ we denote the standard (semantic) entailment relationship; $\models \phi$ means ϕ is a tautology; $\phi \models \psi$ means ψ follows from ϕ. For the purpose of this section we fix a finite number of sentence letters $\mathsf{Q} := \{q_1, \ldots, q_h\}$.

A **choice function** is a function $\pi : N \rightarrow \mathscr{P}(\mathsf{Q})$; here N is again the set of agents, and intuitively $\pi(i)$ provides the information on the choices of agent i. Π is the set of all such functions. Given $Q \subseteq \mathsf{Q}$, ϕ_Q is the formula:

$$\phi_Q := \bigwedge_{q_i \in Q} q_i \;\wedge \bigwedge_{q_i \in (\mathsf{Q} - Q)} \neg q_i$$

If $\phi_{\pi(i)} \models \psi$ then we say that "agent i accepts ψ". The set of all agents that accept $q_j \in \mathsf{Q}$, that is $\{i \in N \mid q_j \in \pi(i)\}$, is denoted by $[\![q_j]\!]_\pi$. More generally, for $\psi \in \mathscr{L}_c$, $[\![\psi]\!]_\pi := \{i \in N \mid \phi_{\pi(i)} \models \psi\}$.

A **social aggregation function (SAF)** is a (possibly partial) function $F : \Pi \rightarrow \mathscr{P}(\mathscr{L}_\mathsf{c})$; $F(\pi)$ denotes the socially accepted sentences of $\mathscr{L}_c$ given π. Both this and the following terminology is simply a restatement of terminology introduced in section 3.1:

Definition 33. *Let $\pi, \pi' \in \Pi$, $\phi, \psi \in \mathscr{L}_c$ be arbitrary. A SAF is said to satisfy:*
universal domain (UD) *iff the domain of F is Π;*
monotonicity (M) *iff whenever $[\![\phi]\!]_\pi \subseteq [\![\phi]\!]_{\pi'}$ then $\phi \in F(\pi) \implies \phi \in F(\pi')$;*
neutrality (N) *iff whenever $[\![\phi]\!]_\pi = [\![\psi]\!]_{\pi'}$ then $\phi \in F(\pi) \iff \psi \in F(\pi')$.*

One can see that there is a nice relation between the framework introduced in this section and that of the previous section on judgement aggregation by observing that (1) the set of all formulae of classical propositional logic satisfies the conditions of being an agenda, and (2) individual judgement sets for this agenda are completely determined by a choice of accepted sentence letters. By our previous results, then, we know that requiring a SAF to respect the three axioms listed above "builds in" the possibility of inconsistent outcomes of the aggregation procedure. However, we remind the reader that this is not our primary concern in this part of the thesis. In fact we now introduce a logic that is perfectly able to handle these kinds of inconsistencies.

4.2 Semantics Based on SAFs

We will be concerned with the following language, whose semantic interpretation will be defined in terms of SAFs. This language $\mathscr{L}_\Box$ is grammatically generated by:

$$\psi ::= \Box\alpha \mid \psi_1 \wedge \psi_2 \mid \neg\psi \mid \bot \qquad \text{with each } \alpha \in \mathscr{L}_c.$$

The interpretation of $\Box\psi$ is that "ψ is collectively accepted". The proposed interpretation of the $\Box$ operator leads us to consider the following natural semantics for the language $\mathscr{L}_\Box$: we interpret the formulae using SAFs and choice functions. The $\Box$ serves to shield the logic of group decisions from the (possibly logically inconsistent) outcome of the aggregation procedure. This gives the language distinct modal flavour, however there are no (iterated) modalities and also no boxless formulae (except $\bot$).

Definition 34. *Let F be a SAF, and π a choice function. The pair (F, π) is called a **$\mathscr{L}_\Box$-model**. Let $\psi, \psi_1, \psi_2 \in \mathscr{L}_\Box$ and $\Psi \subseteq \mathscr{L}_\Box$. We write:*

$$\begin{array}{rll}
 & (F,\pi) \Vdash \Box\phi & \textit{iff } \phi \in \mathscr{L}_c \textit{ and } \phi \in F(\pi); \\
 & (F,\pi) \Vdash \psi_1 \wedge \psi_2 & \textit{iff } (F,\pi) \Vdash \psi_1 \textit{ and } (F,\pi) \Vdash \psi_2; \\
 & (F,\pi) \Vdash \neg\psi & \textit{iff } (F,\pi) \nVdash \psi; \\
 & (F,\pi) \Vdash \bot & \textit{never}, \\
\textit{and:} & F \Vdash \psi & \textit{iff for all } \pi \in \textit{dom}(F), (F,\pi) \Vdash \psi, \\
\textit{and finally:} & F \Vdash \Psi & \textit{iff for all } \psi \in \Psi, F \Vdash \psi.
\end{array}$$

The origin of these ideas is found in Pauly [35],[34], but readers familiar with his work should be warned that our choice of semantics differ in subtle nevertheless important details. Let us briefly dwell on the differences. In Pauly's work, a model for the language $\mathscr{L}_\Box$ is a valuation map $V : \mathscr{L}_\Box \to \{0,1\}$. This map assigns a truth value to each and every formulae of $\mathscr{L}_\Box$. Stated differently, a model is given by a subset of $\mathscr{L}_\Box$, namely precisely those formulae that are true given the valuation V. We might call such a model a **reduced $\mathscr{L}_\Box$-model**, since it abstracts from the interpretation that the truth values of $\mathscr{L}_\Box$ formulae arise through a social choice procedure.

From the perspective of these reduced $\mathscr{L}_\Box$-models, many of our $\mathscr{L}_\Box$-models are indistinguishable. Formally, let us write $(F,\pi) \equiv (F',\pi')$ if and only if for all $\phi \in \mathscr{L}_\Box$, $(F,\pi) \Vdash \phi \iff (F',\pi') \Vdash \phi$. If we denote the equivalence class of an $\mathscr{L}_\Box$-model (F,π) under $\equiv$ by $[(F,\pi)]$ then it is clear that all $(F',\pi') \in [(F,\pi)]$ give rise to the precisely the same reduced $\mathscr{L}_\Box$-model and are thus indistinguishable. In fact, there is a 1-1 relation between the equivalence classes of $\equiv$ and the set of all different possible valuations V.

A social choice theorist will be interested in the properties of SAFs, but within any given equivalence class of reduced models $[(F, \pi)]$, the properties of the many underlying SAFs that generate the same reduced $\mathscr{L}_\Box$-model will generally differ wildly. Hence we have chosen to augment the approach adopted by Pauly by explicitly factoring in the underlying SAFs. The reason is that we would like to link up the business of social choice theorists with that of logicians.

Axiomatisations: Social Choice versus Logic

Let M be a class of $\mathscr{L}_\Box$-models and let S be a theory in the language $\mathscr{L}_\Box$. A logician would say that S **axiomatises** M if M is the class of all $\mathscr{L}_\Box$-models such that $(F, \pi) \in \mathsf{M}$ and $\phi \in S$ implies $(F, \pi) \Vdash \phi$. In this case S uniquely determines M. Similarly, a social choice theorist would say that a collection of axioms A axiomatises a class K of SAFs if $F \in \mathsf{K}$ if and only if F satisfies all axioms in A.

While these two notions of axiomatisations are certainly related, it appears that the social choice theorist is not interested in $\mathscr{L}_\Box$-models *per se*, but rather in the underlying SAFs that "generate" the models.[17] Clearly, some knowledge of properties of the underlying SAFs tells us something about the models they can generate. For instance:

Lemma 35. *Let (F, π) be such that the SAF F is Monotonic and Neutral and satisfies Universal Domain. Let:*

$$RM := \{\Box\phi \to \Box\psi \mid \ \models \phi \to \psi\}. \tag{RM}$$

Then $(F, \pi) \Vdash \chi$ for all $\chi \in RM$.

Proof. Let ϕ, ψ be such that $\phi \to \psi$ is a tautology of classical propositional logic. Suppose $F, \pi \Vdash \Box\phi$. Let $A = [\![\phi]\!]_\pi$ and $B = [\![\psi]\!]_\pi$. Since $\models \phi \to \psi$, it must be the case that $B \supseteq A$. Consider any π' such that $[\![\phi]\!]_{\pi'} = B$. Since F is monotonic, $(F, \pi') \Vdash \Box\phi$. Since F is neutral and $[\![\phi]\!]_{\pi'} = [\![\psi]\!]_\pi$, $(F, \pi') \Vdash \Box\phi$ iff $(F, \pi) \Vdash \Box\psi$. Conclude that $(F, \pi) \Vdash \Box\psi$, as required. ■

Conversely, the models that F can generate might tell us something about the properties of F. Now if $(F, \pi) \Vdash \phi$, we say that ϕ is **satisfied** on the model (F, π). However, to establish properties of the underlying SAFs, we somehow need to abstract from away from those induced by a particular choice function π. An $\mathscr{L}_\Box$ formula ϕ is called **valid** on F if and only if $F \Vdash \phi$. The concept of validity provides exactly the

[17] For this reason, the axiomatisation results of classes of reduced $\mathscr{L}_\Box$-models in Pauly [34] have no straightforward interpretation in terms of the axiomatisation results typically provided by social choice theorists.

abstraction we need, since it is clearly a property of the SAF F itself and not of any particular model based on F.

This notion of validity stems from the semantics of modal logic, in which the same distinction is made between satisfaction of a formula ϕ as a property of models, and validity as a property of some underlying "frame". Therefore we will look at SAFs as a variant of such modal frames. In the balance of this part of the thesis, we will again be concerned with the logic of SAFs that are monotonic and neutral and satisfy universal domain. We restrict attention to these SAFs because there is a very straightforward link between modal logic and the logic of aggregation functions through simple games. $\mathscr{L}_\Box$-models based on such SAFs will be called **simple models**.

4.3 Passing from Simple Games to Social Choice and Vice-Versa

Let $\Omega = (N, W)$ be a simple game. Define:

$$F_\Omega(\pi) := \{\psi \in \mathscr{L}_c \mid \exists A \in W\ \forall i \in A\ \phi_{\pi(i)} \models \psi\},$$

In words, $\psi \in F_\Omega(\pi)$ if there is some winning coalition A of Ω such that every agent $i \in A$ accepts ψ. Clearly F_Ω is a SAF that satisfies M, N, and UD. In fact, we have the following, mirroring the results in this subsection 3.3:

Proposition 36. *The following are equivalent.*
(a) F is a SAF satisfying M, N, UD;
(b) F is a SAF generated by a unique simple game Ω.

The theory of F is the set of formulae $\{\phi \in \mathscr{L}_\Box \mid F \Vdash \phi\}$. Given our previous results, it is unsurprising that some properties of a simple game Ω pass at once to the theory of F_Ω:

Lemma 37. *Let $\Omega = (N, W)$ a simple game.*
(a). For all $\psi \in \mathscr{L}_c$, $F_\Omega \Vdash \Box\psi \to \neg\Box\neg\psi$ iff Ω is proper;
(b). For all $\psi \in \mathscr{L}_c$, $F_\Omega \Vdash \neg\Box\neg\psi \to \Box\psi$ iff Ω is strong.

Proof. (a $\Leftarrow$). Let π be an arbitrary choice function. Let $\psi \in F_\Omega(\pi)$. Then there is $A \in W$ such that $A = [\![\psi]\!]_\pi$. As for any $Q \subseteq \mathsf{Q}$, $\phi_Q \models \psi \iff \phi_Q \not\models \neg\psi$, we have $N - [\![\psi]\!]_\pi = [\![\neg\psi]\!]_\pi$. By (M2a), $N - [\![\psi]\!]_\pi \notin W$. Suppose towards a contradiction that $\neg\psi \in F_\Omega(\pi)$. Then there is $B \in W$ such that every agent $i \in B$ accepts ψ. Clearly $B \subseteq [\![\neg\psi]\!]_\pi$, so by (M1) $[\![\neg\psi]\!]_\pi = N - [\![\psi]\!]_\pi \in W$, a contradiction. Hence $(F, \pi) \Vdash \neg\Box\neg\psi$.
(a $\Rightarrow$). Let ψ be any formula that isn't a tautology or a contradiction. Suppose $F_\Omega \Vdash \Box\psi \to \neg\Box\neg\psi$, and that $A \in W$. Let π be any choice function such that $[\![\psi]\!]_\pi = A$, and so $(F_\Omega, \pi) \Vdash \Box\psi$. Clearly $[\![\neg\psi]\!] = N - A$. Since $F_\Omega \Vdash \Box\psi \to \neg\Box\neg\psi$, we have $(F_\Omega, \pi) \models \neg\Box\neg\psi$, and thus $N - A \notin W$.
(b $\Leftarrow$). Let π be an arbitrary choice function. Suppose $\neg\psi \notin F_\Omega(\pi)$. Then there is no

$A \in W$ such that every agent $i \in A$ accepts $\neg\psi$. In particular $[\![\neg\psi]\!]_\pi \notin W$. But then by (M2b), $N - [\![\neg\psi]\!]_\pi = [\![\neg\neg\psi]\!]_\pi = [\![\psi]\!]_\pi \in W$, and thus $(F_\Omega, \pi) \Vdash \Box\psi$.
(b $\Rightarrow$). Let ψ be any formula that isn't a tautology or a contradiction. Suppose $F_\Omega \Vdash \neg\Box\neg\psi \rightarrow \Box\psi$. Suppose $A \notin W$. Let π be any choice function such that $[\![\neg\psi]\!]_\pi = A$ and hence $\neg\psi \notin F_\Omega(\pi)$. Then $(F_\Omega, \pi) \Vdash \neg\Box\neg\psi$. As $F_\Omega \Vdash \neg\Box\neg\psi \rightarrow \Box\psi$, we find $F_\Omega \models \Box\psi$. So $[\![q_1]\!]_\pi = N - A \in W$. ■

The formula $\Box\phi \rightarrow \neg\Box\neg\phi$ should of course be read as "if ϕ is collectively accepted, then $\neg\phi$ isn't collectively accepted". Hence, as before, the interpretation of the above result is that it shows the important rôle of families of majorities as simple games that are neither too cautious nor too resolute. Intuitively, if Ω is a family of majorities, then F_Ω selects either ψ or $\neg\psi$, and never both.

4.4 Majority Logic

We are now ready to begin a more systematic study of the language of group decisions, or majority logic, that was defined in section 2. It was alluded to above that the language $\mathscr{L}_\Box$ has a distinct modal flavour, and that we might look upon SAFs as close cousins of the modal notion of a "frame". This way of looking at things is justified, at least for M-N-UD-SAFs that concern us in this text, by proposition 36. For instance, observe that simple games allow us to refine the first line in the truth conditions stated in definition 34:

$$(F_{(N,W)}, \pi) \Vdash \Box\phi \quad \text{iff} \quad [\![\phi]\!]_\pi \in W \tag{1}$$

The first aim is to investigate the expressive power of $\mathscr{L}_\Box$. The next subsection looks at invariance results for the language, and we shall see that RK-projection plays a prominent rôle as a morphism between simple models. Thereafter, we expand our view and show how $\mathscr{L}_\Box$ fits into the richer modal logic. Finally, we apply tools from modal logic to arrive at some definability results.

Invariance

RK-projection of simple games was already introduced in section 2. Based on RK-projection, we now define way of creating new simple models out of old that preserves the truth of $\mathscr{L}_\Box$ formulae. The relation $\leq_{\text{RK}}$ can be extended to simple models in a straightforward way.

Definition 38. ***RK-projection of simple models.*** *Let $(F_{(N,W)}, \pi)$ and $(F_{(N',W')}, \pi')$ be simple models. Define the relation $\leq^{\text{M}}_{\text{RK}}$ by:*

$$(F_{(N,W)}, \pi) \leq^{\text{M}}_{\text{RK}} (F_{(N',W')}, \pi') \text{ if and only if there is } f : N' \rightarrow N \text{ s.t.}$$
$$W = f_*(W') \text{ and for all } f(i) \notin \mathscr{D}((N,W)), \pi'(i) = \pi(f(i)).$$

It turns out that this notion of RK-projection is the most natural notion of morphism for simple models. From the perspective of modal logic this does not come as a great surprise, since the construction is akin to the familiar notion of bounded morphism [4]. (Note however that the "dummy clause" allows one to "throw away" information about dummy players, and this has some subtle consequences.) $\mathscr{L}_{\Box}$-truths are invariant under RK-projection:

Lemma 39. *Let* $(F_{(N,W)}, \pi) \leq^{\mathrm{M}}_{\mathrm{RK}} (F_{(N',W')}, \pi')$. *Then for all* $\phi \in \mathscr{L}_{\Box}$, $(F_{(N,W)}, \pi) \Vdash \phi \iff (F_{(N',W')}, \pi') \Vdash \phi$.

Proof. By induction on the complexity of ϕ. ■

We will say that two simple models (F, π) and (F', π') are **isomorphic** if $(F, \pi) \leq^{\mathrm{M}}_{\mathrm{RK}} (F', \pi') \leq^{\mathrm{M}}_{\mathrm{RK}} (F, \pi)$. Clearly, if simple models are isomorphic, they make the same $\mathscr{L}_{\Box}$-formulae true. The converse, however, is false.

Majority Logic as a Fragment of Modal Logic

The language $\mathscr{L}_{\Box}$ is quite plainly a fragment of modal logic, $\mathscr{L}_{\Box\Box}$, which makes use of the grammar:

$$\psi ::= q \mid \neg\psi \mid \psi_1 \wedge \psi_2 \mid \Box\psi \mid \bot \qquad \text{with each } q \in \mathsf{Q}$$

At the same time, the semantics provided by simple models can be seen as a fragment of the standard semantics for monotonic modal logic. Hence we obtain a relation between modal logic and majority logic at the semantic and the syntactic level. This relation is the subject of this subsection. Some familiarity with monotonic modal logic is assumed, refer to Hansen [18] for a thorough introduction. As a brief reminder, in monotonic modal logic formulae are interpreted using neighbourhood semantics:

Definition 40. *A (monotonic)* ***neighbourhood frame (n.f.)*** *is a pair* (S, ν), S *is a nonempty set of states,* $\nu : S \to \mathscr{P}(\mathscr{P}(S))$ *is the neighbourhood function; for each* $s \in S$, $\nu(s)$ *satisfies* (M1). *A* ***neighbourhood model (n.m.)****,* $\mathfrak{M} = (S, \nu, V)$*, is a n.f. paired with a valuation* $V : W \to \mathscr{P}(\mathsf{Q})$.

Formulae of $\mathscr{L}_{\Box\Box}$ are interpreted relative to states, and the semantics of monotonic modal logic will be clear to anyone familiar with normal modal logic, with the possible exception of the modal clause:

$$\mathfrak{M}, s \Vdash \Box\psi \quad \text{iff} \quad \{s \in S \mid \mathfrak{M}, s \Vdash \psi\} \in \nu(w). \tag{2}$$

If a formula ψ is true globally (that is, at all states of a n.m.), we write $\mathfrak{M} \Vdash \psi$. If a formula is valid on a n.f. (i.e., true under all valuations) we write $(S, \nu) \Vdash \psi$.

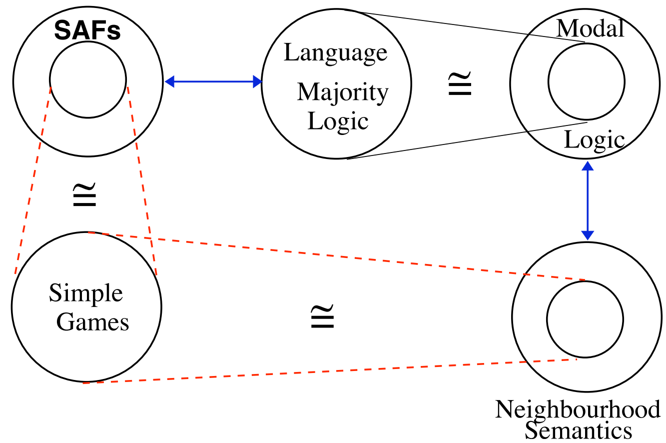


Figure 3: Relations between monotonic modal logic and majority logic

Note that expression (2) contains essentially the same thought as (1) above, except that the set $\nu(s)$ might vary depending on the state we are looking at. A simple model based on a M-N-UD-SAF F and choice function π can be viewed as a n.m.. Under this interpretation, the states of the n.m. become the agents of the corresponding simple game, and we take $V(i) = \pi(i)$ for all $i \in N$. Moreover, the neighbourhood function is a constant function, so that $\nu(i) = W$ for all $i \in N$ (where W is the set of winning coalitions of the simple game that generates F). When we take this perspective, we will denote the corresponding n.m. (or n.f.) simply by (F, π) (or F), and use $\Vdash$ for the truth conditions of both $\mathscr{L}_\Box$ and $\mathscr{L}_{\Box\Box}$ (admittedly with *abuse de langage*). Also from this perspective, an easy induction shows that formulae of $\mathscr{L}_\Box$ have the distinct property that if they are true at *some* state (or agent) in a simple model (F, π), they are true at *all* states. In contrast, note that $\mathscr{L}_{\Box\Box}$ allows us to express some things at the level of individual agents:

$$\text{for all } \psi \in \mathscr{L}_c, \quad (F, \pi), i \Vdash \psi \quad \text{iff} \quad \phi_{\pi(i)} \models \psi.$$

Figure 3 provides an overview of the grand scheme of syntactic and semantic relations.

The language $\mathscr{L}_{\Box\Box}$ can be used to express additional properties of SAFs.

Example 41. *Fix a set N and let $\Omega = (N, \{N\})$. F_Ω is the **consensus-SAF**. It can be shown that $F = F_\Omega$ if and only if F satisfies M, N, and UD and $F \Vdash \Box p \rightarrow p$.*

Hence among M-N-UD-SAFs, $\Box p \rightarrow p$ defines consensus; however, consensus is not expressible by majority logic, since this property is not invariant under adding dummies to Ω, and thus not invariant under RK-projection. We will show next that this is exactly the idea needed to separate $\mathscr{L}_\Box$ from $\mathscr{L}_{\Box\Box}$.

Definition 42. ***RK-Invariance.*** *Let (F, π) and (F, π') be simple models. A for-*

mula $\phi \in \mathscr{L}_{\Box\Box}$ *is* RK*-invariant iff whenever* $(F, \pi), i \Vdash \phi$ *and* $(F', \pi') \leq^{\mathrm{M}}_{\mathrm{RK}} (F, \pi)$*, then* $(F', \pi') \Vdash \phi$*.*

In words, RK-invariance says that local satisfaction of ϕ on (F, π) implies global satisfaction of ϕ in any RK-projection of (F, π) (thus in particular on (F, π) itself!). All $\mathscr{L}_{\Box}$ formulae are RK-invariant, since if $\phi \in \mathscr{L}_{\Box}$ then $(F, \pi), i \Vdash \phi$ implies $(F, \pi) \Vdash \phi$, and furthermore $(F', \pi') \Vdash \phi$ whenever $(F', \pi') \leq^{\mathrm{M}}_{\mathrm{RK}} (F, \pi)$, by lemma 39.

Proposition 43. *Let* $\phi \in \mathscr{L}_{\Box\Box}$*. Then* ϕ *is equivalent to a formula* $\psi \in \mathscr{L}_{\Box}$ *on all simple models if and only if* ϕ *is* RK*-invariant.*

Possibly the proposition can be proved in a syntactic way, e.g. by using reductions to modal normal forms (*à la* Fine [13]). In this text our focus has been firmly on the semantic perspective, and we will seek a proof that is somewhat along the lines of the Van Benthem characterisation result, a corner stone of normal modal logic (see [4]). We need an auxiliary definition and result.

Definition 44. ***Monotonic bisimulation*** *[18, 4.10]. Suppose* $\mathfrak{M} = (S, \nu, V)$ *and* $\mathfrak{M}' = (S', \nu V)$*. Let* $Z \subseteq S \times S'$ *a nonempty relation.* Z *is a* ***bisimulation*** *between* $\mathfrak{M}$ *and* $\mathfrak{M}'$ *if the following three conditions hold: (Prop). If* sZs'*, then* s *and* s' *satisfy the same sentence letters; (Forth). If* sZs' *and* $X \in \nu(s)$*, then there is* $X' \subseteq S'$ *such that* $X' \in \nu'(s')$ *and for all* $s' \in X'$*, there is* $s \in X$ *s.t.* sZs'*; (Back). If* sZs' *and* $X' \in \nu'(s')$*, then there is* $X \subseteq S$ *such that* $X \in \nu(s)$ *and for all* $s \in X$*, there is* $s' \in X'$ *s.t.* sZs'*.*

If Z is a bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$ and sZs', then $\mathfrak{M}, s \Vdash \phi$ if and only if $\mathfrak{M}', s' \Vdash \phi$, for all ϕ in the modal language $\mathscr{L}_{\Box\Box}$ (and thus in $\mathscr{L}_{\Box}$).

Given the finite set of sentence letters Q, consider all formulae build up from literals, conjunction and disjunction. Up to logical equivalence there are only finitely many such formulae, hence we the set Γ that contains all of them (up to logical equivalence) is a finite set. Now define $\Gamma^{\Box}$ as the set:

$$\Gamma^{\Box} := \{\neg\Box\neg\phi \mid \phi \in \Gamma\}.$$

Note that $\Gamma^{\Box}$ is a finite set, and that each $\phi \in \Gamma^{\Box}$ is an $\mathscr{L}_{\Box}$ formula. Let us write $\mathfrak{M} \overset{\Gamma}{\sim} \mathfrak{M}'$ just in case for all $\phi \in \Gamma^{\Box}$, for all states s of $\mathfrak{M}$, and for all states s' of $\mathfrak{M}'$ we have $\mathfrak{M}, s \Vdash \phi \iff \mathfrak{M}', s' \Vdash \phi$. And like before, let us write $\mathfrak{M} \equiv \mathfrak{M}'$ just in case for all $\phi \in \mathscr{L}_{\Box}$, for all states s of $\mathfrak{M}$, and for all states s' of $\mathfrak{M}'$ we have $\mathfrak{M}, s \Vdash \phi \iff \mathfrak{M}', s' \Vdash \phi$. Clearly $\mathfrak{M} \equiv \mathfrak{M}'$ implies $\mathfrak{M} \overset{\Gamma}{\sim} \mathfrak{M}'$.

Lemma 45. ***Collapse of Bisimulation****. Suppose* $\mathfrak{M} \overset{\Gamma}{\sim} \mathfrak{M}'$*. Let* Z *be the relation where* sZs' *if and only if* s *and* s' *satisfy the same sentence letters. Then* Z *is a bisimulation between* $\mathfrak{M}$ *and* $\mathfrak{M}'$*.*

Proof. The proof uses ideas from Hansen [18], Proposition 4.31. Let $\mathfrak{M} = (S, \nu, V)$ and $\mathfrak{M}' = (S', \nu', V')$. (Prop). is clear. (Forth). Suppose sZs' and take $X \in \nu(s)$. We would like to find $X' \in \nu'(s')$ such that $\forall s' \in X'$, there is $s \in X$ for which xZx' holds.

Now towards a contradiction suppose there is no such X'. Then for every $Y \in \nu'(s')$, there is an y_i such that for all $x_j \in X$, it is not true that $x_j Z y_i$. This means y_i and x_j differ in their sentence letters, and there must be literals witnessing this; for instance: $y_i \Vdash \neg q$ and $x_i \nVdash \neg q$. Pick one and denote the literal true at y_i but not at x_j by φ_{ij}. Let Δ_i be the set: $\{\phi_{i'j} \mid i = i'\}$. By construction, for each y_i we have $\mathfrak{M}, y_i \Vdash \bigwedge \Delta_i$. Hence:

$$\mathfrak{M}', s' \Vdash \neg\Box\neg \bigvee_i \bigwedge \Delta_i, \tag{3}$$

$$\text{however, } \mathfrak{M}, s \nVdash \neg\Box\neg \bigvee_i \bigwedge \Delta_i. \tag{4}$$

Now $\neg\Box\neg \bigvee_i \bigwedge \Delta_i$ is equivalent to a formula in $\Gamma^{\Box}$. However, the discrepancy between (3) and (4) contradicts our assumption that $\mathfrak{M} \overset{\Gamma}{\sim} \mathfrak{M}'$. The (Back) clause can be proved in similar fashion. ■

Proof of proposition 43. Left-to-right follows from the invariance results above. For the other direction, we will make use of the fact that the standard translation for monotonic neighbourhood semantics allows us to pass between first order logic and $\mathscr{L}_{\Box\Box}$. We do not spell out the details of the translation here; they can be found in Hansen [18]. For our purpose it suffices that such a translation exists. The standard translation of an $\mathscr{L}_{\Box\Box}$-formula χ is denoted $ST_s(\chi)$ (s is the state it is evaluated at). We use $\models$ for the first order entailment relation, for the purpose of this proof.

Assume ϕ is RK-invariant. Let C be a first order formula expressing that ν is a constant function. Define the set of $\mathscr{L}_{\Box}$-consequences of ϕ on constant models:

$$\mathrm{MLC}(\phi) := \{ST_s(\chi) \mid \chi \in \mathscr{L}_{\Box} \text{ and } \{ST_s(\phi), C\} \models ST_s(\chi)\}.$$

If $\{C\} \cup \mathrm{MLC}(\phi) \models ST_x(\phi)$, then by compactness ϕ is equivalent to a formula $\psi \in \mathscr{L}_{\Box}$ on models satisfying C, hence on simple models. Therefore we will show $\{C\} \cup \mathrm{MLC}(\phi) \models ST_x(\phi)$. So assume that $\mathfrak{M} \models \{C\} \cup \mathrm{MLC}(\phi)[s]$. Since $\mathfrak{M} \models C$, we can view $\mathfrak{M}$ as some simple model (F, π). Say $F = F_{\Omega}$, $\Omega = (N, W)$.

Let $T = \{\forall x ST_x(\xi) \mid F, s \models \xi, \text{ and } \phi \in \mathscr{L}_{\Box}\}$; $\mathfrak{M} \models T$. We claim $T \cup ST_y(\phi)$ is consistent. For suppose not, then by compactness some finite subset T_0 of T is inconsistent with $ST_y(\phi)$, and we have $ST_y(\phi) \to \neg \bigwedge T_0$. Hence $ST_y(\phi) \to \{\exists x \neg ST_x(\xi_1) \vee \cdots \vee \exists x \neg ST_x(\xi_k)\}$. But then $\{C, ST_y(\phi)\} \models \forall x \neg ST_x(\xi_1) \vee \cdots \vee \forall x \neg ST_x(\xi_k)\}$ (using the fact that C forces a constant neighbourhood function, and $\xi_1, \ldots, \xi_k$ are formulae of $\mathscr{L}_{\Box}$). Hence it must be that $\bigvee_{j \in \{1,\ldots,k\}} \neg ST_s(\xi_k) \in \mathrm{MLC}(\phi)$. But this contradicts $T_0 \subseteq T$. So

$T \cup ST_y(\phi)$ is consistent, and hence can be satisfied in some model, say $\mathfrak{N} = (S, \nu, V)$, at some state s^*. Since $\mathfrak{N} \models T$, we know $\mathfrak{N}$ makes exactly the same $\mathscr{L}_\Box$-formulae true as F, and thus $\mathfrak{N} \equiv (F, \pi)$. Now let:

$$D := \{V(s) \mid s \in S \text{ and there is no } i \in N, \pi(i) = V(s)\}.$$

We can add dummies to Ω to account for these all 'missing valuations', and obtain a simple model (F', π'); $(F, \pi) \leq^{\mathrm{M}}_{\mathrm{RK}} (F', \pi')$. Suppose $F' = F'_{N', W'}$. Let $Z \subseteq S \times N'$ be the relation where sZi if and only if s and i satisfy the same sentence letters. By the previous lemma Z is a bisimulation. Moreover, there is a state i^* such that $s^* Z i^*$. Hence $(F', \pi') \Vdash \phi$. By our invariance assumption, for all $x \in N$, $F, \pi \models ST_y(\phi)[x]$—as required. ■

4.5 Axiomatic Social Choice and N.f. Definability

Consider again example 41 above. It illustrates an important conceptual point. In social choice theory, axioms are used to pick out certain classes of social aggregation functions. In modal logic, the notion of frame validity gives a handle on the definability of frame classes. In the present case, the $\mathscr{L}_{\Box\Box}$ formula $\Box p \to p$ picks out the simple games of the form $(N, \{N\})$, which are identified with the consensus-SAFs. Thus modal "frame definability", or in the present framework rather modal "simple game" definability, emerges as the natural logical counterpart to the axiomatic approach to social choice. Modal-like languages give us a precise logical tool to formulate certain kinds of axioms studied in social choice; and it is then natural to ask about the expressive strengths of logical languages: are there limits on their expressive power? How does $\mathscr{L}_\Box$ sit inside $\mathscr{L}_{\Box\Box}$? Results in this and the next subsection provide partial answers to some of such questions. They also underline once more the fundamental importance of the notion of RK-projection.

Frame Theory for $\mathscr{L}_\Box$

We start with giving a precise definition of the notion of "simple game" definability.

Definition 46. *Take* K *a class of M-N-UD-SAFs. Given a set of formulae* $S \subseteq \mathscr{L}_{\Box\Box}$ *we say* K *is* ***definable by*** S *when* $F \in \mathsf{K}$ *if and only if* $F \Vdash S$*. We say the class* K *is* ***closed under RK-projection*** *if the set* $\{\Omega \mid F_\Omega \in \mathsf{K}\}$ *is closed under* RK*-projection.*

Since $\mathscr{L}_\Box \subseteq \mathscr{L}_{\Box\Box}$, the definition above serves for both languages. But for now, we will be interested in $\mathscr{L}_\Box$.

Lemma 47. *Let* K *a class of M-N-UD-SAFs.* K *is definable by a set of* $\mathscr{L}_\Box$*-formulae only if it is closed under* RK*-projection.*

Proof. This follows from the invariance results stated in subsection 4.4. ■

While the above result follows straightforwardly from the invariance results obtained earlier, it leads up to a kind of "impossibility theorem". A **dictatorship** is any SAF F that is based upon a simple game (N, W) with $W = \{X \subseteq N \mid i \in X\}$ for some $i \in N$; and hence $(F, \pi) \Vdash \Box\phi$ if and only if $\phi_{\pi(i)} \models \phi$. Our result states that *non-dictatorship isn't expressible in the language* $\mathscr{L}_\Box$ in any interesting way. Put slightly differently, under modest conditions, the class of all dictatorships sits inside all definable classes of simple games, and thus stacking $\mathscr{L}_\Box$-axioms (expressing properties of simple games) the best one can hope for is to eventually end up pinning down dictatorships. Let us say a class of M-N-UD-SAFs is **degenerate** if it contains only "**trivial**" games of the form $(N, \varnothing)$ and/or "**verum**" games of the form $(N, \mathscr{P}(N))$.

Corollary 48. *Any class* K *of M-N-UD-SAFs that is definable by a set of* $\mathscr{L}_\Box$*-formulae is either degenerate or contains all dictatorships.*

Proof. Suppose K isn't degenerate. Take any $F_\Omega \in \mathsf{K}$ such that $\Omega = (N, W)$ is a trivial nor a verum game. Let Ω' be any dictatorship. By lemma 17, $\Omega' \leq_{\text{RK}} \Omega$. ■

The well known Goldblatt-Thomason theorem for normal modal logics gives necessary and sufficient closure conditions for a first-order definable class of frames to be modally definable (see [4, theorem 3.19]). Improving upon lemma 47, we have a result that gives a Goldblatt-Thomason-style characterisation of $\mathscr{L}_\Box$-definable classes of simple games.

Proposition 49. *Take* K *a class of M-N-UD-SAFs.* K *is definable by a set of* $\mathscr{L}_\Box$ *formulae if and only if it is definable by a set of* $\mathscr{L}_{\Box\Box}$ *formulae and closed under* RK*-projection.*

This result provides some insight in how $\mathscr{L}_\Box$ sits inside $\mathscr{L}_{\Box\Box}$, but obviously raises the question which classes are $\mathscr{L}_{\Box\Box}$-definable. Somewhat surprisingly, the subtle answer to this question does not follow straightforwardly from the Goldblatt-Thomason theorem for monotonic modal logics (as found in e.g. [18, theorem 7.23]). In the next subsection, we investigate the expressivity of $\mathscr{L}_{\Box\Box}$ and will see that we can improve somewhat further on this result, at the expense of some restrictions on the class of frames under scrutiny.

Proof. The left to right direction easily follows from lemma 47. For the other direction, suppose K satisfies the closure condition and is definable by a set of $\mathscr{L}_{\Box\Box}$ formulae S, but not by a set of $\mathscr{L}_\Box$ formulae. We will derive a contradiction.

Consider the $\mathscr{L}_\Box$ theory of K, $\Lambda^{\mathsf{K}}_\Box$. Since K isn't definable by a set of $\mathscr{L}_\Box$ formulae, in particular it isn't definable by $\Lambda^{\mathsf{K}}_\Box$. Hence, there is a simple model (F, π), with $F \notin \mathsf{K}$, such that $(F, \pi) \Vdash \Lambda^{\mathsf{K}}_\Box$ but some state s such that $(F, \pi), s \Vdash \neg\psi$ for some $\psi \in S$.

Consider the set $\Gamma^{\Box} \subseteq \mathscr{L}_{\Box}$ defined earlier. Let Λ^{Γ} be the (finite) set:

$$\Lambda^{\Gamma} := \{\phi \in \mathscr{L}_{\Box} \mid \phi \in \Gamma^{\Box} \text{ and } (F, \pi) \Vdash \phi\} \cup \{\neg\phi \in \mathscr{L}_{\Box} \mid \phi \in \Gamma^{\Box} \text{ and } (F, \pi) \nVdash \phi\}.$$

We claim Λ^{Γ} is satisfiable on some simple game in K. For suppose not, then $\neg \bigwedge \Lambda^{\Gamma} \in \Lambda^{\mathsf{K}}_{\Box}$, which would contradict that $F, \pi \Vdash \Lambda^{\mathsf{K}}_{\Box}$.

So let $(F', \pi') \Vdash \Lambda^{\Gamma}$, with $F' \in \mathsf{K}$. We have established that $F \overset{\Gamma}{\sim} F'$. F' is generated by a simple game $\Omega' = (N', W')$; now consider a dummy extension Ω'' of Ω' with a new agent s', and a corresponding choice function π'' such that $\pi''(i) = \pi'(i)$ for all $i \in N'$, and $\pi''(s') = \pi(s)$. By lemma 45 s' is bisimilar to s. Hence $(F_{\Omega''}, \pi'') \Vdash \neg\psi$. But $\psi \in S$; and $F_{\Omega''} \in \mathsf{K}$ since K is closed under RK-projection; and S defines K—a contradiction. ∎

Frame Theory for $\mathscr{L}_{\Box\Box}$

There are three well known structural operations for normal modal logics that preserve truth, which have analogues in monotonic modal logic: (surjective) bounded morphisms, generated submodels and disjoint unions. To add some perspective to what we are doing in this subsection, we will start by pointing out the explicit connection between (surjective) bounded morphisms and generated submodels and different flavours of RK-projection in the context of simple models. Unfortunately, disjoint unions lack an analogue. The reason is that for neighbourhood models the construction crucially depends on the possibility that the neighbourhood function might vary between states, which is precluded in the context of simple models.

Definition 50. *Let $(F_{(N,W)}, \pi)$ and $(F_{(N',W')}, \pi')$ be simple models.*
A function $f : N \to N'$ is a ***surjective bounded morphism*** *if (i) f is a surjective RK-projection and (ii) $\pi(i) = \pi'(f(i))$ for all $i \in N$.*
$(F_{(N,W)}, \pi)$ is a ***generated submodel*** *of $(F_{(N',W')}, \pi')$ if (N, W) is a dummy contraction of (N', W') and π is just π' restricted to the set N. $(F_{(N,W)}, \pi)$ is the* ***submodel generated by*** *$A \subseteq N'$ if it is the smallest generated submodel such that $A \subseteq N$.*

The constructions above are just restatements of the definitions given for bounded morphisms and generated submodels for monotonic modal logic given in [18, Def. 4.3] and [18, Def. 4.15], restatements that take into account the 'special' structure of simple models. As a consequence, we obtain the following result stemming from monotonic modal logic:

Lemma 51. *(a) Let $(F_{(N,W)}, \pi)$ and $(F_{(N',W')}, \pi')$ be simple models and $f : N \to N'$ a surjective bounded morphism. For all $\phi \in \mathscr{L}_{\Box\Box}$ and all $i \in W$, $(F_{(N,W)}, \pi), i \Vdash \phi$ iff $(F_{(N',W')}, \pi'), f(i) \Vdash \phi$.*
(b) Let $(F_{(N,W)}, \pi)$ be a generated submodel of $(F_{(N',W')}, \pi')$. For all $\phi \in \mathscr{L}_{\Box\Box}$ and all $i \in N$, $(F_{(N,W)}, \pi), i \Vdash \phi$ iff $(F_{(N',W')}, \pi'), i \Vdash \phi$.

For proof, see propositions 4.5 and 4.16 in [18]. Observe how certain kinds of RK-projections on the underlying simple games provide the substrate for these operations on simple models. To obtain the operations at the level of simple games rather than simple models, one deletes the clauses on choice functions from definition 50. Surjective bounded morphisms then collapse to surjective RK-projection, and generated submodels to dummy contractions *sec*.

We have seen that dummy extensions are crucial to separate $\mathscr{L}_{\Box\Box}$ from $\mathscr{L}_{\Box}$. In definition 4 we also introduced the distinction between dummy extension and dummy expansion. Example 41 shows that dummy *extensions* do not preserve the validity of $\mathscr{L}_{\Box\Box}$ formulae. Dummy *expansions*, on the other hand, do.

Lemma 52. *Let $\Omega = (N, W)$ be a simple game with $\mathscr{D}(\Omega) \neq \varnothing$. Let $\Omega' := (N', W')$ be any dummy expansion of Ω. Then for all $\phi \in \mathscr{L}_{\Box\Box}$, $F_\Omega \Vdash \phi$ implies that $F_{\Omega'} \Vdash \phi$.*

Proof. Suppose, to the contrary, that $F_{\Omega'} \nVdash \phi$ while $F_\Omega \Vdash \phi$. There exists a choice function π' and an agent i' such that $(F_{\Omega'}, \pi'), i' \nVdash \phi$. As a consequence of lemma 51, it must be the case that $i' \notin N$. Now $\mathscr{D}(\Omega)$ isn't empty so contains at least one dummy i; and $i \in \mathscr{D}(\Omega')$ since Ω' is a dummy expansion of Ω. Let π'' be the permutation of π' such that $\pi''(i') = \pi'(i)$; $\pi''(i) = \pi'(i')$; and for $j \in N - \{i, i'\}$ to $\pi''(j) = \pi'(j)$. We have $(F_{\Omega', \pi''}), i \nVdash \phi$.

Let π be the restriction of π'' to N. Observe that since Ω is a dummy contraction of Ω', (F_Ω, π) is a generated submodel of $(F_{\Omega'}, \pi'')$. Hence for all $\psi \in \mathscr{L}_{\Box\Box}$, $(F_\Omega, \pi), i \Vdash \psi$ if and only if $(F_{\Omega'}, \pi''), i \Vdash \psi$. Thus in particular $(F_\Omega, \pi''), i \nVdash \phi$, a contradiction. ∎

As an easy corollary of the above results, any $\mathscr{L}_{\Box\Box}$-definable class of simple games must be closed under dummy *expansion* and *contraction*, and under *surjective* RK-projection. All are specific instances of RK-projection.

The preceding analysis shows that certain insights from monotonic modal logic can be easily applied to the logic of simple games, which is, after all, a proper fragment of it, both syntactically and semantically. Because an analogue for the disjoint union construction is lacking, however, it turns out that this fragment is less well behaved than one might expect on an *a priori* basis. Unfortunately, disjoint union functions rather prominently in modal definability results.

We conclude this part of the thesis with a definability result for *finite* simple games. We will show that under a restriction on the size of the set of agents the closure conditions outlined above are not just necessary but also sufficient. For future work it remains to be investigated which modal tools can, and which tools can't be applied when one lifts this limitation. To state our results we will need the following notion:

Definition 53. *Take K and C classes of M-N-UD-SAFs. Given a set of formulae $S \subseteq \mathscr{L}_{\Box\Box}$ we say K is **definable by S relative to** C when for all $F \in \mathsf{C}$, we have $F \in \mathsf{K}$ if* *and only if $F \Vdash S$.*

In what follows we take $\mathsf{C} = \{F_{(N,W)} \mid |N| \leq \log_2 h\}$. The idea behind this restriction is that our language should contain enough sentence letters to simultaneously express everything there is to say about some $F \in \mathsf{C}$. In this case have the following result:

Proposition 54. *Let $\mathsf{K} \subseteq \mathsf{C}$ be a class of SAFs. K is definable relative to C by a set S of $\mathscr{L}_{\Box\Box}$ formulae if and only if it is closed under dummy expansion, dummy contraction, and surjective* RK*-projections.*

Before we move on to its proof, let us point out an immediate corollary of this result, that settles the expressivity of the language $\mathscr{L}_{\Box}$ under the under the same restriction on the size of the set of agents.

Corollary 55. *Let $\mathsf{K} \subseteq \mathsf{C}$ be a class of SAFs. K is definable relative to C by a set S of $\mathscr{L}_{\Box}$ formulae if and only if it is closed under* RK*-projections.*

Proof. The left to right direction follows from observations on stated above. For the other direction, we will show that the $\mathscr{L}_{\Box\Box}$-theory of K, $\Lambda^{\mathsf{K}}_{\Box\Box}$, defines K. To this end, we need to show that if $F \Vdash \Lambda^{\mathsf{K}}_{\Box\Box}$, then $F \in \mathsf{K}$.

So suppose $F \Vdash \Lambda^{\mathsf{K}}_{\Box\Box}$. Without loss of generality, we may assume that the simple game generating F, (N, W) contains at most one dummy. For if (N, W) contains dummies and its dummy contraction with just one dummy can be shown to belong to K then membership of F follows since K is closed under dummy expansions.

We will consider a valuation π that assigns at least one sentence letter q_i to each subset of $\mathscr{P}(N)$. Hence for each $X \subseteq N$, there is a sentence letter q_i such that $(F, \pi), s \Vdash q_i$ if and only if $s \in X$. We will denote this sentence letter by q_X. Note that our condition on C guarantees that we have enough sentence letters available.

Let $d \in N$ be the dummy of (N, W) if it contains one, otherwise d may be chosen arbitrarily from N. Define the set Δ:

$$
\begin{aligned}
\Delta := &\{q_X \mid X \subseteq N \text{ and } (F, \pi), d \Vdash q_X\} \\
&\cup \{\neg q_X \mid X \subseteq N \text{ and } (F, \pi), d \Vdash \neg q_X\} \\
&\cup \{\Box \phi \mid \phi \in \Gamma \text{ and } (F, \pi), d \Vdash \Box \phi\} \\
&\cup \{\neg \Box \phi \mid \phi \in \Gamma \text{ and } (F, \pi), d \Vdash \neg \Box \phi\}
\end{aligned}
$$

Δ is a finite set and we claim it is satisfiable on some simple game in K. For suppose not, then $\neg \bigwedge \Delta \in \Lambda^{\mathsf{K}}_{\Box\Box}$, contradicting that $(F, \pi), d \Vdash \Delta$. So let $(F', \pi'), d' \Vdash \Delta$. We claim there exists a surjective RK-projection from F' to F. From this fact and the closure conditions on K it will follow $F \in \mathsf{K}$.

For $i \in N$, define ψ_i to be the formula:

$$\psi_i := \bigwedge_{\{X \subseteq \mathscr{P}(N) \mid i \in X\}} q_X \wedge \bigwedge_{\{X \subseteq \mathscr{P}(N) \mid i \notin X\}} \neg q_X$$

It is easy to check that $(F, \pi), s \Vdash \psi_i$ if and only $s = i$.

Let i^* be an arbitrary element of N and $f : N' \to N$ be defined thus:

$$f(i') \mapsto \begin{cases} i & \text{only if } F', i' \Vdash \psi_i \\ i^* & \text{otherwise} \end{cases}$$

For each $i \in N$ there is exactly one state at which ψ_i holds, and thus f is well defined.

Claim: f is surjective. Clearly $f(d') = d$. Now let $i \neq d$. We will show that there is $i' \in N'$ such that $f(i') = i$. Since $i \notin \mathscr{D}(\Omega)$, there is some set Z_i such that $Z_i \notin W$ but $Z_i^+ := Z_i \cup \{i\} \in W$.

Subclaim: Let α be the formula $\bigvee_{s \in Z_i^+} \psi_s \wedge \bigwedge_{s \in N - Z_i^+} \neg\psi_s$; then $(F, \pi) \Vdash \Box\alpha$.

It suffices to show that $t \in Z_i^+$ implies $F, t \Vdash \alpha$. This gives $Z_i^+ \subseteq [\![\alpha]\!]_\pi$, and thus $(F, \pi) \Vdash \Box\alpha$. Now, $t \in Z_i^+$ implies $t \neq s$ for all $s \in N - Z_i^+$; this in turn implies $(F, \pi), t \Vdash \neg\psi_s$ for all $s \in N - Z_i^+$. This implies $(F, \pi), t \Vdash \bigwedge_{s \in N - Z_i^+} \neg\psi_s$. Moreover, $(F, \pi), t \Vdash \psi_t$, and hence $(F, \pi), t \Vdash \bigvee_{s \in Z_i^+} \psi_s$. Combining, $(F, \pi), t \Vdash \alpha$, proving our subclaim.

Subclaim: Let β be the formula $\bigvee_{s \in Z_i} \psi_s \wedge \bigwedge_{s \in N - Z_i} \neg\psi_s$. Then $(F, \pi) \Vdash \neg\Box\beta$.

It suffices to show that $(F, \pi), t \Vdash \beta$ implies $t \in Z_i$. In this case $[\![\beta]\!]_\pi \subseteq Z_i$ and thus $[\![\beta]\!]_\pi$ is not a winning coalition. Now, $(F, \pi), t \Vdash \beta$ implies $(F, \pi), t \Vdash \bigvee_{s \in Z_i} \psi_s$ which implies $t \in Z_i$, proving our subclaim.

Now we go on to prove there is i' such that $f(i') = i$. To this end it suffices to show that there is $i' \in N'$ such that $(F', \pi'), i' \Vdash \psi_i$. Suppose towards a contradiction there is no such i'. We know that $(F', \pi'), d' \Vdash \Box\alpha$. Then there is $X \in W'$ such that $t \in X \implies (F', \pi'), t \Vdash \alpha$. In particular, $t \in X$ implies $(F', \pi'), t \Vdash \neg\psi_s$ for all $s \in N - Z_i^+$; and $(F', \pi'), t \Vdash \psi_s$ for some $s \in Z_i^+$. Moreover, we assumed $(F', \pi'), t \nVdash \psi_i$. Hence in fact $t \in X$ implies $(F', \pi'), t \Vdash \neg\psi_s$ for all $s \in N - Z_i$; and $(F', \pi'), t \Vdash \psi_s$ for some $s \in Z_i$. Hence $(F', \pi'), t \Vdash \Box\beta$. But $X \in W$ and so $(F', \pi'), d' \Vdash \Box\beta$ whereas $(F, \pi), d \Vdash \neg\Box\beta$: a contradiction. This proves our claim.

Claim: f is an RK-projection.

We will show: $X \in W \iff f[X] \in W'$. ($\Rightarrow$). Suppose $X \in W$. Then $(F, \pi), d \Vdash \Box \bigvee_{i \in X} \psi_i$. Hence $(F', \pi'), d' \Vdash \Box \bigvee_{i \in X} \psi_i$. Hence the set:

$$X' := \{s' \in N' | (F', \pi'), s' \Vdash \bigvee_{i \in X} \psi_i\} = [\![\bigvee_{i \in X} \psi_i]\!]_{\pi'}$$

is a winning coalition. Now $s' \in X'$ implies $f(s) \in X$ by the definition of f. So $X' \subseteq f[X]$. By monotony, $X' \in W'$ implies $f[X] \in W'$.

($\Leftarrow$). Suppose $X \notin W$. Then $(F, \pi), d \not\Vdash \Box \bigvee_{i \in X} \psi_i$ and hence $(F', \pi'), d' \not\Vdash \Box \bigvee_{i \in X} \psi_i$. Hence the set:

$$X' := \{s' \in N' | (F', \pi'), s' \Vdash \bigvee_{i \in X} \psi_i\} = [\![\bigvee_{i \in X} \psi_i]\!]_{\pi'}$$

isn't a winning coalition. Again $s' \in X'$ implies $f(s) \in X$ by the definition of f. So $X' \subseteq f[X]$. Suppose $f[X]$ is a winning coalition—this will lead to a contradiction. In this case, it must be that X' is a strict subset of $f[X]$; so let $Y = f[X] - X'$. We claim that for each $y \in Y$, $(F', \pi), y \not\Vdash \psi_i$ for any $i \in N$. Suppose not. Since $y \notin X'$, we know $(F', \pi'), y \not\Vdash \psi_i$ for all $i \in X$. Hence $(F', \pi'), y \Vdash \psi_j$ for some $j \in N - X$. Then by definition of f, $f(y) = j$ and thus $f(y) \notin f[X]$, a contradiction.

To complete our argument, for $y \in Y$ let ϕ_y denote the formula:

$$\phi_y := \bigwedge_{\{X | (F', \pi'), y \Vdash q_X\}} q_X \wedge \bigwedge_{\{X | (F', \pi'), y \not\Vdash q_X\}} \neg q_X.$$

Then, as $f[X]$ was a winning coalition:

$$(F', \pi'), d' \Vdash \Box \left(\bigvee_{i \in X'} \psi_i \vee \bigvee_{i \in Y} \phi_i \right).$$

But since by our argument above, for $i \in Y, j \in N$, we have $\phi_i \neq \psi_j$, clearly:

$$(F, \pi), d \not\Vdash \Box \left(\bigvee_{i \in X'} \psi_i \vee \bigvee_{i \in Y} \phi_i \right),$$

giving a contradiction. We conclude $X' = f[X]$ and thus $f[X] \notin W'$. This proves our claim and completes the proof of the proposition. ∎

5 Conclusion

In this thesis we have examined the relationship between logic and social choice theory through simple games. Simple games provide a coherent framework to explore many facets of social choice theory. In fact, simple games correspond in 1-1 fashion to the class of M-N-UD-SAFs, and this has allowed us to explore links between these simple games on the one hand, and aggregation procedures and monotonic modal logic on the other. At the same time, this correspondence also means that a focus on simple games necessarily restricts attention to this class of SAFs. This might be termed a limitation of our work. However, such SAFs have functioned rather prominently in social choice theory, because the conditions imposed by the axioms of Neutrality, Monotonicity and Universal Domain are quite natural in many contexts.

Of these three axioms, perhaps Neutrality is the most controversial, in part because in Arrow's original theorem, a weaker axiom, namely **Independence of Irrelevant Alternatives** is imposed. Geanakoplos [15] and others have shown that, at least for the traditionally studied problem of aggregation of rational preference relations, and given Arrow's other axioms, Neutrality is in fact equivalent to Independence of Irrelevant Alternatives. In the more general judgement aggregation framework, some further conditions on the agenda under scrutiny are needed to establish this equivalence; precise conditions are explored in [10], [11] and [32]. In fact, Blau [5] already proved that Independence of Irrelevant Alternatives is equivalent to various related but apparently stronger axioms. Therefore, a weakening of Neutrality typically does not provide a way out of the "impossibility results": results stating that natural aggregation procedures might not provide logically consistent social opinions. For this reason, we have chosen not to dwell on such weakenings in this thesis.

Apart from consistency, one might be interested in other behavioural properties of SAFs. A specific choice of logical language—majority logic in this thesis—allows us to get a firm logical grip on the axioms that can be formulated within that language, and then to compare the expressive power and relative complexity of different languages for the purpose of axiomatic social choice theory. To this end, one can apply tools from the logician's toolbox to study what properties of SAFs can and can't be defined in the language, and investigate the logical consequences. This has been the main subject of section 4.

One feature that is brought to the fore by this approach is language-dependency. A logic is a formal language that is typically formulated with a specific purpose in mind, and therefore a logical investigation will be language dependent. This dependency can be seen both in the formal dimension, relating to the logical strength of the language relative to other languages, as well as the informal dimension, relating to the kinds of objects the logic language can talk about. Such language dependent investigations can teach us something about the axiomatic approach, since the strength and scope of a language reflects the kind of axioms that be formulated within it. For instance, the language $\mathscr{L}_\Box$ expresses facts about the logical structure of socially accepted statements. Therefore axioms in the language $\mathscr{L}_\Box$ can be used to express certain consistency constraints on social opinions.

While in section 4 we have not been concerned explicitly with the impossibility results obtained in social choice theory, they emerge quite easily in this framework. To avoid that logical inconsistencies arise in the aggregation process, we would like a SAF to respect the rules of classical logic. In "axiomatic" terms, what we need is the SAF to validate the formula $\Box p \leftrightarrow \neg\Box\neg p$, and in addition we want $\Box$ to distributive: $\Box p \wedge \Box q \leftrightarrow \Box(p \wedge q)$. These $\mathscr{L}_\Box$-formulae force the underlying simple game to be a family of majorities, and a filter, respectively. The only simple games that satisfy these properties correspond to the ultrafilters (indeed the formulae *define* this class of M-N-UD-SAFs); and hence the impossibility results emerge. This analysis shows that to avoid the impossibility results, one will have to give up either one of the primitive logical connectives $\wedge$ or $\neg$. Distributivity is required to preserve the $\wedge$-connective, and the ultraproperty gives the law of the excluded middle.

With these observations in mind, we would like to end this thesis with some thoughts on the remarkable robustness of Arrow's theorem. The same "dictatorial"-style impossibility results have emerged in many contexts and domains over the past 60 years, and hence scores of years after Arrow's discovery, his theorem is still found at the very core of social choice theory. In our view, the results of the two parts of this thesis provide two clues why this might be so. A first clue is provided by the observation that the constraints on SAFs which tend to produce dictatorial procedures—that the underlying simple games satisfy the ultraproperty and are closed under intersections—appear to be intimately related to the primitive logical connectives $\wedge$ or $\neg$. In this light it is a bit unsurprising that any approach that imposes some kind of logical constraint on social opinions tends to produce dictatorial SAFs.[18]

A second clue is provided by our result that the language $\mathscr{L}_\Box$ cannot express non-dictatorship, so that axioms formulated in the language tend to narrow down the corresponding class of SAFs about the dictatorial SAFs. Perhaps many "natural" or "interesting" axioms suffer from the same defect, and thus lead to constraints that actually push the class of SAFs that satisfy them *towards* dictatorships. Given the structure of the language $\mathscr{L}_\Box$, it is clear that axioms that express consistency constraints on social opinions will have this property. However, to make this observation more precise, or more general, what is needed is a clear mathematical conceptualisation of what it means for an axiom to be "natural" or "interesting". We leave such exercises for future work,

[18]Social choice theorists often argue that the classical impossibility results are produced by a clash between Independence of Irrelevant Alternatives, which specifies how an aggregation procedure should behave on pairs, and Transitivity, which specifies how it should behave on triples. Our argument also shows that this is only part of the story.

but hope that our analysis of $\mathscr{L}_\Box$ provides a nice starting point for the investigation of richer languages.

ACKNOWLEDGEMENT: I am indebted to my mentor Benedikt Löwe and my supervisor Eric Pacuit for guiding me through the breathtaking universe of logic during the past two years. AMSTERDAM, MARCH 2007.

References

[1] ARROW, K. Le principe de rationalité dans les décisions collectives. *Économie Appliquée 5*, 4 (1952), 501–84.

[2] ARROW, K. *Social Choice and Individual Values*, 2nd ed. No. 12 in Cowles Foundation Monographs. New York etc.: Wiley, 1963.

[3] BATTEAU, P., BLIN, J.-M., AND MONJARDET, B. Stability of aggregation procedures, ultrafilters, and simple games. *Econometrica 49*, 2 (1981), 527–34.

[4] BLACKBURN, P., DE RIJKE, M., AND VENEMA, Y. *Modal Logic*. Cambridge: Cambridge University Press, 2001.

[5] BLAU, J. H. Arrow's theorem with weak independence. *Economica 38*, 152 (1971), 413–20.

[6] DANIËLS, T. R., AND PACUIT, E. A general approach to aggregation problems. mimeo.

[7] DAVEY, B. A., AND PRIESTLEY, H. A. *Introduction to Lattices and Order*, 2nd ed. Cambridge University Press, Cambridge, 2002.

[8] DIETRICH, F. A generalized model of judgement aggregation. *Social Choice and Welfare* (Forthcoming).

[9] DIETRICH, F., AND LIST, C. Judgement aggregation without full rationality. Unpublished version dated August 2006.

[10] DIETRICH, F., AND LIST, C. Arrow's theorem in judgement aggregation. *Social Choice and Welfare* (Forthcoming).

[11] DOKOW, E., AND HOLZMAN, R. Aggregation of binary evaluations. Tech. rep., Technion Israel Institute of Technology, 2005.

[12] ECKERT, D., AND PIGOZZI, G. Belief merging, judgment aggregation, and some belief merging, judgment aggregation, and some links with social choice theory. In

Belief Change in Rational Agents: Perspectives from Artificial Intelligence, Philosophy, and Economics (2005), no. 05321 in Dagstuhl Seminar Proceedings, Internationales Begegnungs- und Forschungszentrum fuer Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI).

[13] Fine, K. Normal forms in modal logic. *Notre Dame journal of formal logic 16*, 2 (1975), 229–37.

[14] Gärdenfors, P. A representation theorem for voting with logical consequences. *Economics and Philosophy 22* (2006), 181–90.

[15] Geanakoplos, J. Three brief proofs of Arrow's impossibility theorem. Cowles Foundation Discussion Papers 1123R3, Cowles Foundation, 1996.

[16] Guenin, L. M. The set theoretic ambit of Arrow's theorem. *Synthese 126*, 3 (2001), 443–72.

[17] Guilbaud, G. Th. Les théories de l'intérêt général et le problème logique de l'agrégation. *Économie Appliquée 5*, 4 (1952), 501–84.

[18] Hansen, H. H. Monotonic modal logics. ILLC Preprint Series PP-2003-24, Institute of Logic, Language and Computation, Universiteit van Amsterdam, 2003.

[19] Hodges, W. *A Shorter Model Theory.* Cambridge etc.: Cambridge University Press, 1997.

[20] Isbell, J. A class of simple games. *Duke Mathematical Journal 25* (1958), 423–39.

[21] Jech, T. *Set Theory*, third millenion edition ed. Springer Monographs in Mathematics. Berlin etc.: Springer-Verlag, 2006.

[22] Kirman, A., and Sondermann, D. Arrow's theorem, many agents and invisible dictators. *Journal of Economic Theory 5* (1972), 267–77.

[23] Kornhauser, L. A., and Sager, L. G. Unpacking the court. *Yale Law Review 82* (1986).

[24] List, C., and Pettit, P. Aggregating sets of judgements: An impossibility result. *Economics and Philosophy 18* (2002), 89–110.

[25] Mas-Colell, A., Whinston, M. D., and Green, J. R. *Microeconomic Theory.* Oxford: Oxford University Press, 1995.

[26] Maskin, E. S. Majority rule, social welfare functions, and game forms. In *Choice, Welfare, and Development (essays in honor of Amartya Sen)*, K. Basu, P. Pattanaik, and K. Suzumura, Eds. Oxford: Oxford University Press, 1995, ch. 6, pp. 100–109.

[27] May, K. O. A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica 20*, 4 (1952), 680–4.

[28] Mihara, H. R. Arrow's theorem and Turing computability. *Economic Theory 10*, 2 (1997), 257–76.

[29] Monjardet, B. An axiomatic theory of tournament aggregation. *Mathematics of Operations Research 3*, 4 (1978), 334–51.

[30] Monjardet, B. De Condorcet à Arrow via Guilbaud, Nakamura et les «jeux simples». *Mathematics and Social Sciences 41*, 163 (2004), 5–32.

[31] Monjardet, B. Social choice theory and the "Centre de Mathématique Sociale": Some historical notes. *Social Choice and Welfare 25*, 2 (December 2005), 433–56.

[32] Nehring, K., and Puppe, C. Consistent judgement aggregation: A characterization. Tech. rep., University of Karlsruhe, 2005.

[33] Pacuit, E., and Salame, S. Majority logic. Unpublished version dated 8th September 2005.

[34] Pauly, M. Axiomatising collective judgment sets in a minimal logical language. Unpublished version dated 2nd June 2006.

[35] Pauly, M. On the role of language in social choice theory. Unpublished version dated 7th November 2005.

[36] Pauly, M., and van Hees, M. Logical constraints on judgement aggregation. *Journal of Philosophical Logic* (forthcoming).

[37] Rubinstein, A. The single profile analogues to multi profile theorems: Mathematical logic's approach. *International Economic Review 25*, 3 (October 1984), 719–30.

[38] Schoenmakers, W. A problem in knowledge acquisition. *SIGART Newsletter* (1986).

[39] Sen, A. *Handbook of Mathematical Economics*, vol. 3. Amsterdam etc.: North-Holland, 1986, ch. Social Choice Theory, pp. 1073–181.

[40] Shapley, L. S. Simple games: an outline of the descriptive theory. *Behavioral Science 7* (1962), 59–66.

[41] Suppes, P. The pre-history of Kenneth Arrow's social choice and individual values. *Social Choice and Welfare 25*, 2–3 (2005), 319–26.

[42] Taylor, A. D., and Zwicker, W. S. *Simple Games. Desirability Relations, Trading and Pseudoweightings.* Princeton: Princeton University Press, 1999.

[43] Von Neumann, J., and Morgenstern, O. *Theory of Games and Economic Behavior.* Princeton: Princeton University Press, 1944.

[44] Wilson, R. The game-theoretic structure of Arrow's general possibility theorem. *Journal of Economic Theory 5* (1972), 14–20.