Social Choice and the Logic of Simple Games

Tijmen R. Daniëls

Abstract

From the perspective of the minimal majority logic proposed by Pauly [12], we investigate the relation between axiomatic social choice theory, the logic of simple games, and neighbourhood semantics. We discuss the importance of the Rudin-Keisler ordering in this context and provide a simple characterisation of the monotonic modal fragment that corresponds to the logic of simple games based on this ordering. Finally we discuss its relevance for axiomatic social choice theory.

1 Introduction

Social choice theory is concerned with questions on how a group of agents can decide as a collective in a way that reflects the individual opinions of those involved. The rich history of the subject can be traced back more than two centuries, to eighteenth century thinkers such as Jeremy Bentham, Jean-Charles de Borda, and especially the Marquis de Condorcet. For many years the work done by these thinkers laid dormant. Then social choice suddenly picked up steam in the 1950's, when the economist Kenneth Arrow used observations originally made by Condorcet to prove a striking result, *viz.*, that it is impossible to aggregate rational preference relations into a collective (or social) rational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 axioms, or 'democratic' desiderata [1]. Many similar results followed in its wake.

Some recent work on social choice has revolved around **judgement aggregation** (a non-exhaustive list includes [3], [8], [10], [13]). This work is concerned with the question of aggregating a collection of sentences in of a formal logical language, in a logically consistent way, and by a method reflecting the individual views of a group of 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 virtually all results on preference aggregation have their counterparts in this newer context. And indeed, the interest in judgement aggregation was spawned initially by the discovery of an Arrow style impossibility result (List and Pettit, [8]).

In our view, however, there are at least two 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 very explicitly brings out the connection between logic and social choice theory. A link between social choice and logic has always been present—on occasions Kenneth Arrow has recounted that his interest in applying axiomatic methods to social choice had sprung from exposure to the mathematics of Gödel and Tarski. But the new context has inspired logicians to investigate higher-order questions about social choice using formal methods. One promising way to go about is to define a formal language which can formalise certain behavioural properties of aggregation procedures. Recently, Pauly [12] has provided a modal-flavoured logic of collective decision making that does just that.

This paper is in this more recent logical tradition. It is not so much concerned with impossibility results *per se*, but rather with placing social choice in the context of methods familiar to logicians. We will be working with a formal language of collective decision making in the tradition of Pauly [12], defined in section 2. Instead of studying the language in isolation, we will make use of the artillery provided by monotonic modal logic and simple games. The importance of the latter to understanding social choice has been stressed by e.g. Monjardet [9]. Section 3 discusses such simple games in some depth; we present a generalisation of Monjardet's results to the logic of collective decision making and look at the Rudin-Keisler ordering on simple games. In section 4 we relate this perspective to monotonic modal logic. We work towards a simple characterisation that shows how the logic of collective decision procedures fits into the larger modal picture. We conclude with some implications for the axiomatic method: application of standard methods gives insight into what classes of social aggregation procedures can be defined in simple modal languages.

1.1 Preliminaries

We will define a basic language $\mathscr{L}_c$ that is just 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 text we follow the standard conventions for bracketing and use the abbrevations $\rightarrow$, $\leftrightarrow$, $\vee$. By $\models$ we denote the standard (semantic) entailment relationship; $\models \varphi$ means φ is a tautology; $\varphi \models \psi$ means ψ follows from φ.

For the purpose of this paper we fix a finite number of sentence letters $\mathsf{Q} := \{q_1, \ldots, q_h\}$. N is the set of agents—whenever we assume N finite we will state this explicitly. A **choice function** is a function $\pi : N \rightarrow \mathscr{P}(\mathsf{Q})$; 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:

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

If $\varphi_{\pi(i)} \models \psi$ 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 $\psi \in \mathscr{L}_c$, $[\![\psi]\!]_\pi := \{i \in N \mid \varphi_{\pi(i)} \models \psi\}$.

A **social aggregation function (SAF)** is a (possibly partial) function $F : \Pi \to \mathscr{P}(\mathscr{L}_{\mathsf{c}})$; $F(\pi)$ denotes the socially accepted sentences of $\mathscr{L}_c$ given π. The following terminology is standard:

Definition 1 *Let $\pi, \pi' \in \Pi$, $\varphi, \psi \in \mathscr{L}_c$ be arbitrary. A SAF is said to satisfy:*
universal domain (UD) *iff the domain of F is Π;*
monotonicity (M) *iff whenever $[\![\varphi]\!]_\pi \subseteq [\![\varphi]\!]_{\pi'}$ then $\varphi \in F(\pi) \implies \varphi \in F(\pi')$;*
neutrality (N) *iff whenever $[\![\varphi]\!]_\pi = [\![\psi]\!]_{\pi'}$ then $\varphi \in F(\pi) \iff \psi \in F(\pi')$.*

2 Semantics Based on SAFs

Our point of departure will be 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. The origin of these ideas is Pauly [12], but readers familiar with that paper should be warned that the present semantics differ in details: Pauly's models assign truth values directly to the formulae of $\mathscr{L}_\Box$.

Definition 2 *Let F be a SAF, and π a choice function in the domain of F. The pair (F, π) is called a **model**. Let $\psi, \psi_1, \psi_2 \in \mathscr{L}_\Box$ and $\Psi \subseteq \mathscr{L}_\Box$. We write:*

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

Now consider:

$$RE := \{\Box\varphi \leftrightarrow \Box\psi \mid \ \models \varphi \leftrightarrow \psi\} \qquad (\mathsf{RE})$$
$$RM := \{\Box\varphi \to \Box\psi \mid \ \models \varphi \to \psi\} \qquad (\mathsf{RM})$$

It may be verified that the following holds (see also Pauly [12]):

Lemma 3 *If F satisfies* N*, then $F \Vdash RE$. If F satisfies* N *and* M*, then $F \Vdash RE \cup RM$.*

In the balance of this paper, we will be concerned with the logic of SAFs that are monotonic and neutral and satisfy universal domain. Models based on such SAFs will be called **simple models**.

3 Simple Games

Perhaps the most familiar and natural aggregation procedure is simple majority voting. Simple games provide a generalised interpretation of the notion of a "majority". Certainly, 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 is the basic intuition underlying simple games, formulated by Von Neumann and Morgenstern [11], and 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 (or, in game theoretic parlance, the **winning coalitions** of N). Then W is closed under supersets:

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

Hence by a **simple game** we mean a pair (N, W), where N is a nonempty set of agents and $W \subseteq \mathscr{P}(N)$ satisfies condition (M1). A simple game is **finite** if N is a finite set. 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 say, all majorities are **strict**. On the other hand (M2b) expresses that A is a majority whenever its complement isn't. In a historically important paper by G. Th. Guilbaud [5], the proper strong simple games were called **families of majorities**, and we will stick to this terminology below.[1]

A player $i \in N$ 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 its dummy players by $\mathscr{D}(\Omega)$.

[1] In fact, the simple games envisioned by Von Neumann and Morgernstern were both proper and strong. They investigated various properties of such games, including issues of computational complexity.

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

In this subsection we narrow down the relation between simple games and monotonic, neutral and universal domain SAFs to a 1-1 correspondence. These results expand on observations made by Monjardet [9] on preference aggregation. Let $\Omega = (N, W)$ be a simple game. Define:

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

In words, $\psi \in F(\pi)$ 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. Some properties of simple games pass at once to the resulting aggregation function.

Lemma 4 *Let $\psi \in \mathscr{L}_c$ and $\Omega = (N, W)$ a simple game.*
(a). If Ω is proper, then $F_\Omega \models \Box\psi \to \neg\Box\neg\psi$;
(b). If Ω is strong, then $F_\Omega \models \neg\Box\neg\psi \to \Box\psi$.

Proof. Let π be an arbitrary choice function. (a). Let $\psi \in F_\Omega(\pi)$. Then there is $A \in W$ such that every agent $i \in A$ accepts ψ. So $A \subseteq [\![\psi]\!]_\pi$. By (M1), $[\![\psi]\!]_\pi \in W$. As $\varphi_Q \models \psi \iff \varphi_Q \not\models \neg\psi$, 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 $(F, \pi) \Vdash \neg\Box\neg\psi$. (b). 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 $(F, \pi) \Vdash \Box\psi$. ■

One interpretation of the above result is that it shows the important rôle of families of majorities as simple games that are neither too conservative nor too resolute. Intuitively, if Ω is a family of majorities, then F_Ω selects either ψ or $\neg\psi$, and never both. These two horns are expressed by the following schemes:

$$D := \{\Box\varphi \to \neg\Box\neg\varphi \mid \varphi \in \mathscr{L}_c\} \tag{D}$$

$$Dc := \{\neg\Box\neg\varphi \to \Box\varphi \mid \varphi \in \mathscr{L}_c\} \tag{Dc}$$

We say that a SAF F is **generated by a simple game** if there is a simple game Ω such that $F = F_\Omega$.

Proposition 5 *Fix a set of agents N. The following are equivalent.*
(a) F is a SAF satisfying M, N, UD;
(b) F is a SAF generated by a simple game Ω.
Moreover, $F \Vdash D$ iff Ω is proper, and $F \Vdash Dc$ iff Ω is strong.

Proof. (a $\implies$ b). Call a set A ψ-decisive iff $\psi \in F(\psi)$ whenever $[\![\psi]\!]_\pi = A$. If F is neutral, then A is ψ-decisive if and only if there exists $\pi \in \Pi$ such that $[\![\psi]\!]_\pi = A$ and $\psi \in F(\pi)$. Let $W(\psi)$ the family of ψ-decisive sets. By monotony, if A contains a ψ-decisive set, then A is a ψ-decisive set, so W satisfies (M1). By neutrality, for all $\psi, \varphi \in \mathscr{L}_c$, $W(\varphi) = W(\psi) =: W$. So (N, W) is a simple game, and it is straightforward to verify F is generated by (N, W).

If $F \Vdash D$ then Ω is proper: Suppose $F \Vdash D$, and that $A \in W$. Let π be any choice function such that $[\![q_1]\!]_\pi = A$, and so $(F, \pi) \Vdash \Box q_1$. Clearly $[\![\neg q_1]\!] = N - A$. By D, $F \models \neg\Box\neg q_1$, and thus $N - A \notin W$. If $F \Vdash Dc$ then Ω is strong: Suppose $F \Vdash Dc$. Suppose $A \notin W$. Let π be any choice function such that $[\![\neg q_1]\!]_\pi = A$ and $\neg q_1 \notin F(\pi)$. Then $(F, \pi) \Vdash \neg\Box\neg q_1$. By Dc, $F \models \Box q_1$. So $[\![q_1]\!]_\pi = N - A \in W$.

The other halves of the claims follow by lemma 4. ■

3.2 The Rudin-Keisler Ordering

The **Rudin-Keisler (RK) ordering** was introduced by Rudin as an ordering of ultrafilters (see Jech [7]). Taylor and Zwicker [14] observe that this ordering has a natural interpretation when applied to simple games.[2] Formally, if N and M are two sets of agents, and $\Omega = (N, W)$ is a simple game and f is a map from N to M, $f_*(W)$ 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 preimage of A (that is: $\{i \in N \mid f(i) \in A\}$).

RK-*ordering and bloc formation.* When applied to simple games, the game $(M, f_*(W))$ is obtained intuitively by considering the players of Ω identified by f to vote as a bloc. The upshot of this is that any outcome arrived at in $(M, f_*(W))$ can be arrived at in Ω by letting these players vote *en bloc* in this manner.

The following definition of the Rudin-Keisler ordering differs from the one given by Taylor and Zwicker [14] and from the one familiar from the literature on ultrafilters in that we do not require f to be a surjection.

Definition 6 *We say that $\Omega = (N, W)$ is **RK-below** $\Omega' = (N', W')$, iff there exists a map f such that $W = f_*(W')$; in this case we write $\Omega \leq_{\mathrm{RK}} \Omega'$. Games Ω and Ω' are called **isomorphic** if $\Omega \leq_{\mathrm{RK}} \Omega' \leq_{\mathrm{RK}} \Omega$. We will write $\Omega \leq_{\mathrm{RK}}^{\mathrm{SUR}} \Omega'$ if there exists a* surjection *f such that $\Omega = f_*(\Omega')$. Finally, we say an* RK-*projection is **finite** if both N and N' are finite sets.*

If $\Omega \leq_{\mathrm{RK}} \Omega'$ then Ω is called an RK-projection of Ω'. It is not hard to see that the relations $\leq_{\mathrm{RK}}$ and $\leq_{\mathrm{RK}}^{\mathrm{SUR}}$ are transitive and reflexive (and hence pre-orderings) and that $\leq_{\mathrm{RK}}^{\mathrm{SUR}} \subset \leq_{\mathrm{RK}}$. Properties preserved by RK-projection include monotony, properness, and strongness.

Lemma 7 *If $\Omega \subseteq \mathscr{P}(N \cup A)$ is obtained from $\Omega' \subseteq \mathscr{P}(N)$ by adding a set of dummy players A, then Ω is isomorphic to Ω'.*

[2] In fact one does not even need to demand the monotony condition (M1) of the families of sets under consideration—the ordering also makes sense for arbitrary subsets of $\mathscr{P}(N)$.

The analogous claim for $\leq_{\text{RK}}^{\text{SUR}}$ quite obviously fails; which explains our choice of $\leq_{\text{RK}}$ as the default.[3] In fact, it is quite easy to see that if the projection function $f : N \to M$ isn't surjective, then the set $M - \text{ran}(f)$ will consist of dummy players. Hence any RK-projection may be decomposed in a surjective projection and an operation that adds dummies.

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. In fact, we will take a look at 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 5. For instance, observe that simple games allow us to refine the first line in the truth conditions stated in definition 2:

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

The 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.

4.1 Invariance Results

In this section we define two ways of creating new simple models out of old that preserve the truth of $\mathscr{L}_\Box$ formulae. The first two of these constructions stem from the game-theoretical literature on simple games and thus have a natural interpretation outside the logical framework considered in this text [14].

Definition 8 *Let $\Omega = (N, W)$ and $\Omega' = (N', W')$. The* ***product game*** $\Omega \otimes \Omega$ *is given by:*

$$(N \cup N', \{X \subseteq \mathscr{P}(N \cup N') \mid X \cap N \in W \text{ and } X \cap N' \in W'\})$$

The ***bicameral meet*** $\Omega \sqcap \Omega'$ *is the special case where N and N' are disjoint sets.*

The terminology "bicameral meet" comes from the idea that N and N' are two distinct "chambers", and a proposal has to pass both of these chambers to become accepted [14].

[3] Taylor and Zwicker [14] point out the possibility of dropping the surjectivity condition on f in this context.

Lemma 9 *Suppose* $\Omega = (N, W)$ *and* $\Omega' = (N', W')$ *are simple games such that* N *and* N' *are disjoint. Let* π *and* π' *be choice functions such that* $dom(\pi) = N$ *and* $dom(\pi') = N'$, *and let* π'' *be the choice function such that* $dom(\pi'') = N \cup N'$, *and* $\pi''(i) = \pi(i)$ *if* $i \in N$, *and* $\pi''(i) = \pi'(i)$ *if* $i \in N'$. *Let* $\varphi \in \mathscr{L}_\Box$ *and suppose* $F_\Omega, \pi \Vdash \varphi$ *and* $F_{\Omega'}, \pi' \Vdash \varphi$. *Then* $F_{\Omega \sqcap \Omega'}, \pi'' \Vdash \varphi$.

Proof. By induction on the complexity of φ. ■

RK-projection of simple games was already introduced in the previous section.

Definition 10 ***RK-projection of simple models.*** *The relation* $\leq_{\mathrm{RK}}$ *can be extended to simple models as follows. Let* $(F_{(N,W)}, \pi)$ *and* $(F_{(N',W')}, \pi')$ *be models. Define the relation* $\leq_{\mathrm{RK}}^{\mathrm{M}}$ *by:*

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

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 [2]. (Note however that the dummy clause allows one to "throw away" information about certain players, and this has some subtle consequences.) $\mathscr{L}_\Box$-truths are invariant under RK-projection:

Lemma 11 *Let* $(F_{(N,W)}, \pi) \leq_{\mathrm{RK}}^{\mathrm{M}} (F_{(N',W')}, \pi')$ *and* f *such that* $W = f_*(W')$ *and for all* $i \notin \mathscr{D}((N, W)), \pi(f(i)) = \pi(i)$. *Then for all* $\varphi \in \mathscr{L}_\Box$, $(F_{(N,W)}, \pi) \Vdash \varphi \iff (F_{(N',W')}, \pi') \Vdash \varphi$.

Proof. By induction on the complexity of φ. ■

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

The final construction introduced here is inspired by the notion of ultraproducts known from modal logic, rather than by game theory. Let $\{(N_i, W_i)\}_{i \in I}$ be a family of simple games such that the sets N_i are disjoint. Let U be an ultrafilter over I; U may be thought of as the collection of "large subsets" of I.

Definition 12 ***Generalised Meet.*** $\bigsqcap_U (N_i, W_i)$ *is the simple game* (N, W) *such that:*

$$N = \bigcup_{i \in I} N_i, \qquad \text{and} \qquad X \in W \iff \{i \in I \mid X \cap N_i \in W_i\} \in U.$$

A $\varphi \in \mathscr{L}_\Box$ is true in $\bigsqcap_U (N_i, W_i)$ iff it is in a "large set" of underlying models:

Lemma 13 *Let* $\{\pi_i : N_i \to \mathscr{P}(\mathsf{Q})\}_{i\in I}$ *be a collection of choice functions, and let* $\pi : \bigcup_{i\in I} N_i \to \mathscr{P}(\mathsf{Q})$ *be the choice function such that* $\pi(j)$ *is just* $\pi_i(j)$*. For all* $\varphi \in \mathscr{L}_\Box$*,* $\prod_U \Omega_i \Vdash \varphi \iff \{i \in I | \Omega_i \Vdash \varphi\} \in U$*.*

Proof. By induction on the complexity of φ. ■

4.2 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 [6] for a thorough introduction. As a brief reminder, in monotonic modal logic formulae are interpreted using neighbourhood semantics:

Definition 14 *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 $(S, \nu) \Vdash \psi$.

Note that expression (2) contains essentially the same thought as (1) above. A simple model based on a simple game $\Omega = (W, N)$ and choice function π can be viewed as a n.m. where $\nu(i) = W$ and $V(i) = \pi(i)$ for all $i \in N$. For this reason (admittedly with *abuse de langage*) 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 $\mathscr{L}_{\Box\Box}$. 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.

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

Example 15 *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* $F \Vdash \Box p \to p$*.*

Hence among M-N-UD-SAFs, $\Box p \to p$ 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 16 ***RK-Invariance.*** *Let (F,π) and (F,π') be simple models. A formula $\varphi \in \mathscr{L}_{\Box\Box}$ is* RK*-invariant iff whenever $(F,\pi), i \Vdash \varphi$ and $(F,\pi) \leq^{\mathrm{M}}_{\mathrm{RK}} (F',\pi')$, then there is a state (or agent) i' in the model (F',π') such that $(F',\pi'), i' \Vdash \varphi$. In words, satisfaction of φ is preserved under* RK*-projection.*

Proposition 17 *Let $\varphi \in \mathscr{L}_{\Box\Box}$. 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 [4]). In this text our focus has been firmly on the semantic perspective, and we will seek a proof along the lines of the Van Benthem characterisation result, a corner stone of normal modal logic (see [2]). We need an auxiliary definition and result.

Definition 18 ***Monotonic bisimulation*** *[6, 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 $\mathfrak{M}, s \Vdash \varphi$ if and only if $\mathfrak{M}', s' \Vdash \varphi$, for all φ in the modal language $\mathscr{L}_{\Box\Box}$ (and thus in $\mathscr{L}_\Box$).

Let us write $\mathfrak{M} \equiv \mathfrak{M}'$ just in case for all $\varphi \in \mathscr{L}_\Box$, for all states s of $\mathfrak{M}$, and for all states s' of $\mathfrak{M}'$ we have $\mathfrak{M}, s \Vdash \varphi \iff \mathfrak{M}', s' \Vdash \varphi$.

Lemma 19 ***Collapse of Bisimulation****. Suppose $\mathfrak{M} \equiv \mathfrak{M}'$. 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 [6], 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: $\{\varphi_{i'j} \mid i = i'\}$. By construction, for each y_i we have $\mathfrak{M}, y_i \Vdash \bigwedge \Delta_i$. Note that Δ_i is a finite set, since there are only finitely many literals given our assumption on Q. Hence:

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

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

Since there are only finitely many literals, there can be only finitely many different sets Δ_i. Hence without loss of generality, we may assume any disjunction over a conjunction of them is finite; and thus $\neg \bigvee_i \bigwedge \Delta_i \in \mathscr{L}_c$ since it is a finite formula build from propositions, $\wedge$, $\vee$, $\neg$. Hence $\neg\Box\neg \bigvee_i \bigwedge \Delta_i \in \mathscr{L}_\Box$. Clearly, the discrepancy between (3) and (4) contradicts the fact that $\mathfrak{M} \equiv \mathfrak{M}'$. The (Back) clause can be proved in similar fashion. ■

Proof of proposition 17. Left-to-right follows from the invariance results above. As 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 $\mathscr{L}_{\Box\Box}$, see again [6] for details. 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:

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

If $\{C\} \cup \mathrm{MLC}(\varphi) \models ST_x(\varphi)$, 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 $\{C\} \cup \mathrm{MLC}(\varphi) \models ST_x(\varphi)$. Assume that $\mathfrak{M} \models \{C\} \cup \mathrm{MLC}(\varphi)[s]$. We can view $\mathfrak{M}$ as some simple model (F, π). Say $F = F_\Omega$, $\Omega = (N, W)$.

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

4.3 Axiomatic Social Choice and N.f. Definability

Consider again example 15 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, frame validity gives a handle on the definability of frame classes. The $\mathscr{L}_{\Box\Box}$ formula $\Box p \to p$ picks out the simple game $(N, \{N\})$, which is identified with the consensus-SAF. Thus "frame definability", or in the present framework rather "simple game" definability, is the natural logical counterpart to the axiomatic approach to social choice. Modal-like languages gives 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 $\mathscr{L}_{\Box\Box}$? The next two results provides partial answers to some of such questions. They also underline once more the fundamental importance of the notion of RK-projection.

Definition 20 *Let* K *a class of* $\mathsf{M\text{-}N\text{-}UD\text{-}SAFs}$*. We say* K *is closed under* RK*-projection if the set* $\{\Omega \mid F_\Omega \in \mathsf{K}\}$ *is closed under* RK*-projection. Similarly for bicameral meet, etc.*

Proposition 21 *Let* K *a class of* $\mathsf{M\text{-}N\text{-}UD\text{-}SAFs}$*.* K *is definable by a set of* $\mathscr{L}_{\Box}$*-formulae only if it is closed under* RK*-projections and bicameral meet.*

Proof. This follows from the invariance results stated in subsection 4.1. ■

Proposition 22 *Let* K *be a class of* $\mathsf{M\text{-}N\text{-}UD\text{-}SAFs}$ *that is definable by a set of* $\mathscr{L}_{\Box\Box}$ *formulae and that is closed under generalised meet. Then* K *is definable by a set of* $\mathscr{L}_{\Box}$ *formulae if and only if it is closed under* RK*-projection.*

Proof. Suppose K is definable by an $\mathscr{L}_{\Box\Box}$ theory S. We will show the $\mathscr{L}_{\Box}$ theory of K, $\Lambda^{\mathsf{K}}_{\Box}$, defines K, along the lines of a fairly standard argument from modal logic [2]. Suppose the contrary. Then there exists a simple model $\mathfrak{M}$, whose underlying simple game isn't in K, such that $\mathfrak{M} \Vdash \Lambda^{\mathsf{K}}_{\Box}$ but for some state s, $\mathfrak{M}, s \Vdash \neg\psi$, where $\psi \in S$. Let $\Lambda^{\mathsf{M}}_{\Box}$ be the $\mathscr{L}_{\Box}$ theory of $\mathfrak{M}$. Every finite subset of $\Lambda^{\mathsf{M}}_{\Box}$ is satisfiable in some model (Ω, π) in K—for suppose not, then there is a finite subset $F \subseteq \Lambda^{\mathsf{M}}_{\Box}$, $\neg \bigwedge F \in \Lambda^{\mathsf{K}}_{\Box}$, but this contradicts $\mathfrak{M} \Vdash \Lambda^{\mathsf{K}}_{\Box}$.

Define an index set I such that $I = \{F \subseteq \Lambda^{\mathsf{M}}_{\Box} \mid F \text{ is finite}\}$. For each $i \in I$ there is a simple model Ω_i, π_i such that $\Omega_i \Vdash i$. Because K is closed under RK-projection, we may take these models disjoint. For each $\varphi \in S$, let $\hat{\varphi}$ the set of all $i \in I$ that contain $\hat{\varphi}$. The set $\{\hat{\varphi} \mid \varphi \in \S\}$ has the finite intersection property and thus can be extended to an ultrafilter U. Now let $\pi : \bigcup_{i \in I} N_i \to \mathscr{P}(\mathsf{Q})$ be the choice function such that $\pi(j)$ is just $\pi_i(j)$. Then $\prod_U \Omega_i, \pi \Vdash \Lambda^{\mathsf{M}}_{\Box}$. This is true, since for each $i \in \hat{\varphi}$, we have $\varphi \in i$, and hence $\Omega_i, \pi_i \Vdash \varphi$. Therefore $\{i \in I \mid \Omega_i \Vdash \varphi\} \supseteq \hat{\varphi} \in U$ and thus by lemma 13, $\prod_U \Omega_i, \pi \Vdash \varphi$.

Given this model $\prod_U \Omega_i =: \mathfrak{M}^*$, by the closure conditions of K, $\Lambda^{\mathsf{M}}_{\Box}$ is satisfiable on a simple game in K. Now $\mathfrak{M}^* \equiv \mathfrak{M}$. Like in the proof of proposition 17, we may add dummies to obtain a model $\mathfrak{M}^{**}$ and a state s^{**} bisimilar to the state s of $\mathfrak{M}$. It follows $\mathfrak{M}^{**}, s^{**} \Vdash \neg\varphi$ and hence $\mathfrak{M}^{**} \nVdash S$. But $\mathfrak{M}^{*} \leq^{\mathrm{M}}_{\mathrm{RK}} \mathfrak{M}^{**}$, and thus the underlying simple game is in K, a contradiction. ■

5 Concluding Remarks

The main theme of this text is that—from a logical point of view—the axiomatic approach to social choice (which should be distinguished from the *logical* idea of providing axiomatisations) quite naturally corresponds to the investigation of definability results in an appropriately chosen logical language. Let us conclude with two remarks that put this message into some wider perspective.

The first remark is illustrative in nature. While we have not been concerned explicitly with the impossibility results obtained in social choice theory, they emerge quite easily from our 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 strong and proper, and closed under finite intersections. It is well known that 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.

In addition, one might be interested in other behavioural properties of SAFs. A precise choice of logical language—majority logic in this text—allows us to get a firm logical grip on the axioms that can be formulated within a 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 matter of this text. It is worth to point out a related technical observation. We have argued that the semantics for the language $\mathscr{L}_\Box$ somehow sit as a fragment inside the more well-known neighbourhood semantics, and have made good use of this fact too. However, it turns out that the fragment is less well behaved than one might expect on an *a priori* basis. When comparing the structural truth preserving operations set out in section 4.1 with those familiar from modal logic, they appear closely related. What seems to be lacking, however, is an analogue for the disjoint union construction, which functions rather prominently in modal definability results. For future work it remains to be investigated which modal tools can, and which tools can't be applied under this limitation.

References

[1] ARROW, K. *Social Choice and Individual Values*, 2nd ed. No. 12 in Cowles

Foundation Monographs. New York etc.: Wiley, 1963.

[2] Blackburn, P., de Rijke, M., and Venema, Y. *Modal Logic*. Cambridge: Cambridge University Press, 2001.

[3] Dietrich, F., and List, C. Arrow's theorem in judgement aggregation. *Social Choice and Welfare* (forthcoming).

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

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

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

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

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

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

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

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

[12] Pauly, M. Axiomatising judgement aggregation procedures in a minimal logical language. Unpublished version dated 30th March 2005.

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

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

Tijmen R. Daniëls
Amsterdam School of Economics, and Tinbergen Institute,
Universiteit van Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands
Email: tijmen.daniels@uva.nl