The birth of social choice theory from the spirit of mathematical logic: Arrow's theorem as a model-theoretic preservation result

Daniel Eckert and Frederik Herzberg

Logical Models of Group Decision Making (ESSLLI 2013)
August 2013, Düsseldorf

Connections between logic and social choice

Two sources

- Recent Interest of computer science in voting rules (e.g. from an algorithmic point of view) -> necessity for a formal language to represent social choice procedures
- Judgment aggregation: recent generalisation of classical Arrovian social choice from the aggregation of preferences to the aggregation of arbitraty information in some logical language -> necessity for a formal language to reason about the processing of these inputs
- Many different approaches in judgment aggregation! for a survey see e.g. List/Puppe 2009

The contribution of model theory

- Natural approach: Model theory (see e.g. Bell and Slomson 1969) is the study of the relation between (especially relational!!) structures and sentences that hold true in them.
- Recent work by Herzberg and Eckert has proposed a unified framework for aggregation theory (including judgment aggregation) based on the aggregation of model-theoretic structures, thus extending Lauwers and Van Liedekerke's (1995) model-theoretic analysis of preference aggregation. This model-theoretic framework for aggregation theory conceives of an aggregation rule as a map $f : \mathrm{dom}(f) \to \Omega$ with $\mathrm{dom}(f) \subseteq \Omega^I$, wherein I is the electorate and Ω is the collection of all models of some fixed universal theory T (in a first-order language $\mathcal{L}$) with a fixed domain A. This map thus assigns to any profile of models of T an $\mathcal{L}$-structure that is also a model of T.

Thus, in model-theoretic terms, an aggregation rule is equivalent to an operation on a product of models of some theory T that guarantees that the outcome of this operation is again a model of T, i.e. that all the properties of the factor models described by the theory T are preserved. The fact that this is typically not the case for a direct product consisting in a profile of preference orderings lies at the heart of the problem of preference aggregation since Condorcet's paradox about the possibly cyclical outcome of majority voting. This framework is sufficiently general to cover both preference and propositional judgment aggregation: For instance, preference aggregation corresponds to the special case where $\mathcal{L}$ has one binary relation R, T is the theory of weak orders, and A is a set of alternatives; propositional judgment aggregation corresponds to the special case where $\mathcal{L}$ has a unary operator (the belief operator) and A is the agenda. In this model-theoretic approach to aggregation theory, basic (im)possibility theorems from preference aggregation and judgment follow directly from general (im)possibility theorems about the aggregation of first-order model-theoretic structures.

The fundamental observation in the model-theoretic analysis of aggregation is that the preservation of certain properties of the individual factor models requires that the outcome be some reduction of the direct product taken over a family of subsets of the electorate. Once this observation has been made, the proof of characterisations of aggregation functions (in the guise of (im)possibility theorems) only requires relatively basic facts from model theory, such as the construction of reduced products, ultraproducts, Łoś's theorem, and the characterisations of filters and ultrafilters on finite sets. Dictatorship then immediately follows in the finite case, if this family is required to be an ultrafilter, because in this case an ultrafilter is the collection of all supersets of some singleton, - the dictator.

Arrow's theorem as a model-theoretic preservation result

- a model-theoretic approach is not only consistent with Arrow's original research program
- his dictatorship result is a model-theoretic preservation result "avant la lettre", a historical significance that was explicitly recognized by Hodges (2000) in his account of the history of model theory.
- Roughly speaking, this significance consists in the formulation of the problem of the aggregation of preference relations as a typical model-theoretic preservation problem, i.e. as the problem of the preservation of the properties of the individual factor models under product formation, a core problem in the subsequent literature on model theory in the 60s and 70s (see e.g. Chang and Keisler).
- The application of model-theoretic results to preference aggregation can already be found in an old unpublished paper by Brown 1975

- From a methodological point of view, Arrow's seminal 1951 monograph *Social Choice and Individual Values* is rightly famous for its introduction of the axiomatic analysis of binary relations into economics and welfare economics in particular.
- The context of justification of this approach to the modelling of social welfare is the so-called ordinalist revolution of the 1930s, which put into question the measurability and, a fortiori, the interpersonal comparison of utilities.
- But its context of discovery is Arrow's exposure as a student to the work of the famous logician Alfred Tarski, in particular to the algebra of relations in the 1940s.

Textual evidence

- Arrow explicitly motivates the formal framework of binary relations used for the representation of preferences by its familiarity "in mathematics and particularly in symbolic logic" (Arrow, 1963, p. 11), referring to Tarski's famous *Introduction to Logic and the Methodology of the Deductive Sciences*, 1941, which he had proofread as a student.
- More generally, Arrow's analysis of the problem of preference aggregation can be read as an application of the deductive method exposed in Tarski's textbook.
- Central to Tarski's concept of a deductive theory is not only its derivation from a set of axioms, but the concept of a model of a theory obtained by an interpretation of its terms that makes all the axioms (and thus the theory derived from them) true.
- The latter can be seen as the conceptual intuition underlying the further development of model theory as well as of its significance for the epistemological analysis of those social sciences that can be counted among the formal sciences, like theoretical economics.

Another source of inspiration: Karl Menger's semantics of deontic logic

The construction of various types of products with the help of families of sets on some index set would later play a central role in model theory (e.g. in Łoś's 1954 fundamental theorem on ultraproducts), Arrow's analysis of collective decision problems in terms of families of winning coalitions can be traced back to another, "semantical" logical strand in the research program of the mathematization of economics. It was the mathematician Karl Menger who in 1934 first introduced families of subsets of individuals into the logical analysis of norms, semantically conceiving a norm as the set of individuals accepting it.

This approach was then explicitly propagated by Morgenstern in his programmatic paper *Logistics and the Social Science* 1936 as a model for the application of formal analysis to the social sciences in general and to economics in particular. In this light, the analysis of games in terms of families of winning coalitions in von Neumann and Morgenstern's foundational *Theory of Games and Economic Behavior* 1944, to which Arrow often refers, can be considered a significant step in this logical strand in the mathematization of economics.
Thus Arrow's seminal monograph is located at the confluence of two logical strands, Tarski's model-theoretic approach to the methodology of the deductive sciences and Menger's logical semantics of norms in terms of families of subsets of individuals.

Arrow's theorem as a model-theoretic preservation result

A is interpreted as a set of alternatives and T is the theory of weak orders, which is expressed by the universal sentences:
(i) $\forall x \forall y \; R(x,y) \vee R(y,x)$ (completeness, Axiom I in Arrow 1963) and
(ii) $\forall x \forall y \forall z \; R(x,y) \wedge R(y,z) \rightarrow R(x,z)$ (transitivity, Axiom II in Arrow 1963).
Denote by Ω the set of all models of T and by I the (possibly infinite) set of individuals.
A social welfare function is a map f whose domain $\mathrm{dom}(f)$ is contained in Ω^I and whose range is contained in Ω. Under the traditional assumption of universal domain, a social welfare function is then a mapping $f : \Omega^I \rightarrow \Omega$, which assigns to each profile of weak orders a weak order as a social preference. The very definition of a social welfare function, thus, does already imply the requirement of the preservation of the first-order properties of preference relations under product formation.

Analysis of social welfare functions in terms of families of winning coalitions

The following proposition establishes the link between the independence property and the analysis of collective decision problems in terms of families of winning coalitions.

Proposition

A social welfare function $f : \mathrm{dom}(f) \to \Omega$ *satisfies independence of irrelevant alternatives if and only if for any pair of alternatives* $x, y \in A$ *there exists a family of winning coalitions* $\mathcal{W}^f_{(x,y)} \subset 2^I$ *such that for any profile* $\underline{\mathfrak{A}} \in \mathrm{dom}(f)$

$$f(\underline{\mathfrak{A}}) \models R(x,y) \Leftrightarrow \{i \in I : \mathfrak{A}_i \models R(x,y)\} \in \mathcal{W}^f_{(x,y)}$$

Further Arrovian properties

Definition

A social welfare function $f : \Omega^I \to \Omega$ which satisfies independence of irrelevant alternatives is **weakly Paretian**, if for any pair of alternatives $x, y \in A$

$$\varnothing \notin \mathcal{W}^f_{(x,y)}$$

Definition

A social welfare function $f : \mathrm{dom}(f) \to \Omega$ is called **Arrovian** if and only if it has universal domain ($\mathrm{dom}(f) = \Omega^I$), is weakly Paretian and satisfies independence of irrelevant alternatives.

Dictatorship and the social structure

Similarly, the property of non-dictatorship can be characterized via sets of winning coalitions.

Definition

An Arrovoan social welfare function $f : \Omega^I \to \Omega$ is **non-dictatorial**, if there does not exist an individual $k \in I$ such that for all alternatives $x, y \in A$,

$$\mathcal{W}^f_{(x,y)} = \{S \subseteq I : k \in S\}.$$

The implication of preservation of completeness

Lemma

(Strongness) *Let $f : \Omega^I \to \Omega$ be an Arrovian social welfare function (and suppose $\#A \geq 2$). Then for any pair of distinct alternatives $x, y \in A$ and any coalition $U \in 2^I$*

$$U \notin \mathcal{W}^f_{(x,y)} \Rightarrow I \setminus U \in \mathcal{W}^f_{(y,x)}.$$

Proof.

Let $x, y \in A$ with $x \neq y$ and $U \notin \mathcal{W}^f_{(x,y)}$. Since f is a social welfare function with universal domain, we can construct a profile $\underline{\mathfrak{A}} \in \mathrm{dom}(f)$ such that
(a) for all $i \in I$, $\mathfrak{A}_i \models \neg R(x,y) \vee \neg R(y,x)$ (completeness of the negated order), and
(b) $\{i \in I : \mathfrak{A}_i \models R(x,y)\} = U$.
Then, on the one hand $I \setminus U = \{i \in I : \mathfrak{A}_i \not\models R(x,y)\} = \{i \in I : \mathfrak{A}_i \models \neg R(x,y)\} = \{i \in I : \mathfrak{A}_i \models R(y,x)\}$, because our choice of $\underline{\mathfrak{A}}$ and completeness imply $\mathfrak{A}_i \models (\neg R(x,y) \leftrightarrow R(y,x))$ for all $i \in I$ ("$\rightarrow$" by completeness, "$\leftarrow$" by (a)).
On the other hand, by the assumption $U \notin \mathcal{W}^f_{(x,y)}$, we may deduce $f(\underline{\mathfrak{A}}) \not\models R(x,y)$, which by completeness (of the social preference ordering) yields $f(\underline{\mathfrak{A}}) \models R(y,x)$.
Combining this, we conclude $I \setminus U \in \mathcal{W}^f_{(y,x)}$. □

The implication of preservation of transitivity

Lemma

(Monotonicity Lemma) *Let $f : \Omega^I \to \Omega$ be an Arrovian social welfare function (and suppose $\#A \geq 3$). Then for any triple of distinct alternatives $x, y, z \in A$, any winning coalitions $U \in \mathcal{W}^f_{(x,y)}$ and $V \in \mathcal{W}^f_{(y,z)}$, $W \in \mathcal{W}^f_{(x,z)}$ for all $W \supseteq U \cap V$.*

Proof.

Since f is a social welfare function with universal domain, we can construct a profile $\underline{\mathfrak{A}} \in \Omega^I = \mathrm{dom}(f)$ such that
(a) $\{i \in I : \mathfrak{A}_i \models R(x,y)\} = U$,
(b) $\{i \in I : \mathfrak{A}_i \models R(y,z)\} = V$, and
(c) $\{i \in I : \mathfrak{A}_i \models R(x,z)\} = W$.
(This is possible due to the assumption of $W \supseteq U \cap V$ and x, y, z being distinct.)
By (a), (b) and the decisiveness of U, V, $f(\underline{\mathfrak{A}}) \models R(x,y) \wedge R(y,z)$ and hence, by transitivity, $f(\underline{\mathfrak{A}}) \models R(x,z)$. Thus, by independence, $\{i \in I : \mathfrak{A}_i \models R(x,z)\} \in \mathcal{W}^f_{(x,z)}$, whence by (c), $W \in \mathcal{W}^f_{(x,z)}$. □

Model theoretic significance

Simple proof of a generalization of Arrow's theorem which establishes its relation to the ultraproduct construction in model theory by showing that an Arrovian social welfare function is equivalent to the reduction of a direct product of preference relations over an ultrafilter on the set of individuals.
Recall that a *filter* on the set I is a family $\mathcal{W} \subset 2^I$ such that
(F1) $\mathcal{W} \neq \varnothing$ and $\varnothing \notin \mathcal{W}$ (non-triviality)
(F2) $U \cap V \in \mathcal{W}$ for all $U, V \in \mathcal{W}$ (finite intersection closure)
(F3) $V \in \mathcal{W}$ whenever $V \supseteq U$ for some $U \in \mathcal{W}$ (superset closure).
A filter is an *ultrafilter* on I if for any $U \subseteq I$ either $U \in \mathcal{W}$ or $I \setminus U \in \mathcal{W}$.
An ultrafilter $\mathcal{W}$ on I is *principal* if and only if there exists some $k \in I$ such that $\mathcal{W} = \{U \subseteq I : k \in I\}$.
The reduction of a direct product $\underline{\mathfrak{A}}$ over an ultrafilter $\mathcal{W}$ is known as an *ultraproduct* and is denoted by $\underline{\mathfrak{A}}/\mathcal{W}$.

Theorem

Let $f : \Omega^I \to \Omega$ be an Arrovian social welfare function. Then there exists an ultrafilter $\mathcal{W} \subset 2^I$ such that
(i) for any profile $\underline{\mathfrak{A}} \in \Omega^I$ and for all pairs of alternatives $x, y \in A$, $f(\underline{\mathfrak{A}}) \models R(x, y)$ if and only if $\{i \in I : \mathfrak{A}_i \models R(x, y)\} \in \mathcal{W}$, and
(ii) for any profile $\underline{\mathfrak{A}} \in \Omega^I$ and for all pairs of alternatives $x, y \in A$ $f(\underline{\mathfrak{A}}) \models R(x, y)$ if and only if $\underline{\mathfrak{A}}/\mathcal{W} \models R(x, y)$.
In particular, if I is finite, then there is no non-dictatorial Arrovian social welfare function.

Some lemmas for the proof

Lemma

(Contagion Lemma) *Let $f : \Omega^I \to \Omega$ be an Arrovian social welfare function. Then for any two pairs of (possibly nondistinct) alternatives $a, b \in A$ and $x, y \in A$, $\mathcal{W}^f_{(x,y)} = \mathcal{W}^f_{(a,b)}$*

Proof.

Let $a, b, x, y \in A$ and $U \in \mathcal{W}^f_{(x,y)}$. Because of universal domain, we can construct a profile $\underline{\mathfrak{A}} \in \Omega$ such that (a) for all $i \in I$, $\mathfrak{A}_i \models R(a,x) \wedge R(y,b) \wedge R(x,a) \wedge R(b,y)$ and (b) $\{i \in I : \mathfrak{A}_i \models R(x,y)\} = U$.
By transitivity, for all $i \in I$, $\mathfrak{A}_i \models (R(a,b) \leftrightarrow R(x,y))$, and hence $\{i \in I : \mathfrak{A}_i \models R(a,b)\} = U$. By the Pareto principle, $f(\underline{\mathfrak{A}}) \models R(a,x) \wedge R(y,b) \wedge R(x,a) \wedge R(b,y)$ and then by transitivity $f(\underline{\mathfrak{A}}) \models (R(a,b) \leftrightarrow R(x,y))$. However, $f(\underline{\mathfrak{A}}) \models R(x,y)$ due to $\{i \in I : \mathfrak{A}_i \models R(x,y)\} = U \in \mathcal{W}^f_{(x,y)}$. Hence, $f(\underline{\mathfrak{A}}) \models R(a,b)$ and thus $U \in \mathcal{W}^f$

Some lemmas ctd.

This neutrality property immediately strenghtens independence to a property known as systematicity in the literature on judgment aggregation:

Proposition

Let $f : \Omega^I \to \Omega$ be an Arrovian social welfare function. Then f is **systematic**, *i.e. for all $x, y \in A$*

$$\mathcal{W}^f_{(x,y)} = \bigcup_{a,b \in A} \mathcal{W}^f_{(a,b)} = \bigcap_{a,b \in A} \mathcal{W}^f_{(a,b)}$$

In view of this equality, we may henceforth suppress the subscript of $\mathcal{W}^f$. Note that the family of winning coalitions inherits the strongness property of any of the $\mathcal{W}^f_{(x,y)}$.

With these results, the proof of the theorem follows almost immediately.

Proof.

Let $\mathcal{W}$ be the family $\mathcal{W}^f$ of winning coalitions. We verify (i) and (ii) in the Theorem, as follows:
(i) Non-triviality (F1) follows directly from the weak Pareto property combined with the strongness property (which ensures $I \in \mathcal{W}$), while intersection (F2) and superset closure (F3) follow from the Monotonicity Lemma. Moreover, given that $\mathcal{W}$ is a filter, the strongness property implies that it is an ultrafilter.
(ii) Follows directly from part (i) and the (elementary) atomic case of Łoś's theorem. Łoś's theorem is the central theorem on ultraproducts. It asserts in particular that for any profile $\underline{\mathfrak{A}} \in \Omega^I$ and any sentence φ, $\underline{\mathfrak{A}}/\mathcal{W} \models \varphi$ if and only if $\{i \in I : \mathfrak{A}_i \models \varphi\} \in \mathcal{W}$. In our proof, we only need this result for atomic φ, viz. for every $\underline{\mathfrak{A}} \in \Omega^I$ and all $x, y \in A$,

$$\underline{\mathfrak{A}}/\mathcal{W} \models R(x,y) \Leftrightarrow \{i \in I : \mathfrak{A}_i \models R(x,y)\} \in \mathcal{W},$$

which is an immediate consequence of the definition of an ultraproduct. □

Dictatorship, finally

Finally, let I be finite, and suppose, for a contradiction, f were a non-dictatorial Arrovian social welfare function. The finiteness of I implies, by a well-known lemma from Boolean algebra, that $\mathcal{W}$ is principal. Hence in light of (i), there is some individual $k \in I$ such that for all $\underline{\mathfrak{A}} \in \Omega^I$ and all $x, y \in A$, $f(\underline{\mathfrak{A}}) \models R(x, y)$ if and only if $\mathfrak{A}_k \models R(x, y)$. Such an individual k is a dictator, contradiction.

Conclusion

According to Arrow's theorem, it is the requirement of the preservation of the first-order properties of the individual preference relations by an Arrovian social welfare function which establishes the equivalence of the latter with the model-theoretic construction later known as ultraproduct, i.e. the reduction of the direct product over an ultrafilter on the index set of the individuals. A typical preservation problem thus lies at the origin of the development of Arrovian social theory. As dictatorship is just a consequence of the ultrafilter structure of the family of winning coalitions on a finite set of individuals, preservation problems can be seen to lie at the heart of impossibility results in aggregation theory.

- Arrow, K.J. (1963), Social Choice and Individual Values. 2nd ed. Wiley
- Bell, J.L:, Slomson, A.B. (1969), Models and Ultraproducts. An Introduction, North Holland
- Brown, D.J. (1975), Collective Rationality, Technical Report.
- Chang, C.C., Keisler, H.J. (1990). Model Theory, North Holland
- Herzberg, F., Eckert, D. (2012), The model-theoretic approach to aggregation: Impossibility results for finite and infinte electorates, Mathematical Social Sciences 64: 41–47
- Hodges, W.(2000), Model theory. Technical Report, Queen Mary, Univ. of London
- Lauwers, L., Van Liedekerke, L., (1995) Ultraproducts and aggregation, Journal of Mathematical Economics 24: 217-237.


List, C., Puppe, C. (2009), Judgment aggregation, in: Anand, P., Pattanaik, P.K., Puppe, C. eds., Rational and Social Choice: An Overview of New Foundations and Applications. Oxford University Press, Chapter 19.


Menger, K., Moral (1934), Wille und Weltgestaltung. Grundlegung zur Logik der Sitten, Springer


Morgenstern, O. (1936), Logistics and the Social Sciences, Zeitschrift für Nationalökonomie, 7(1):1-24


Neumann, J.v., Morgenstern, O. (1944), Theory of Games and Economic Behavior, Princeton University Press.


Tarski, A. (1941), Introduction to Logic and the Methodology of the Deductive Sciences. Oxford University Press.