Computational Social Choice 2024

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

[http://www.illc.uva.nl/~ulle/teaching/comsoc/2024/]

Plan for Today

References to “logic” in classical SCT are mostly about the axiomatic method, which is logic-like in spirit but doesn’t make use of a formal language with an associated semantics and proof theory.

Today is about *logic for social choice* in a more technical sense: embedding parts of the theory of social choice into a logical system.

Two natural approaches:

- Start from a given logic and explore what we can model
- Start from a given social choice scenario and design a logic for it

We will use *classical first-order logic* to illustrate the first approach, and a tailor-made *modal logic* to illustrate the second.

U. Endriss. Logic and Social Choice Theory. In A. Gupta and J. van Benthem (eds.), *Logic and Philosophy Today*, College Publications, 2011.

Logical Minimalism

But first: why model social choice problems in logic at all?

Besides offering a *deeper understanding* and besides sometimes having direct *practical use* ($\hookrightarrow$ SAT), there also are philosophical arguments.

Pauly (2008) argues for *formal minimalism:*

> When considering an axiom in SCT, besides its *normative appeal* and its *logical strength*, we should also take into account the *expressivity of the language* needed to define it. *Less is better*.

This perspective allows us, for instance, to investigate whether a given rule *can be axiomatised* at all, given constraints on language.

M. Pauly. On the Role of Language in Social Choice Theory. *Synthese*, 2008.

Propositional Logic

The SAT approach requires us to model social choice scenarios in logic. The fact that this works so well for the simplest of all logics, namely propositional logic, actually is somewhat surprising.

Exercise: *What's the main reason why propositional logic was enough?*

Social Welfare Functions

For the main part of today's lecture, we will focus on *Arrow's Theorem* in its original formulation for *social welfare functions:*

$$F : \mathcal{L}(A)^n \to \mathcal{L}(A)$$

Arrow's Theorem says that there exists no social welfare function (SWF) for $m \geqslant 3$ alternatives that is Paretian, IIA, and nondicatorial.

- *Paretian:* If everyone ranks x above y, then so does the collective.
- *IIA:* The relative collective ranking of x and y depends only on the relative individual rankings of x and y.
- *Nondicatorial:* No individual can dictate the collective ranking.

Observe how this mirrors our earlier formulation for SCFs.

First Approach: First-Order Logic

When considering to model a social choice scenario in logic, it makes sense to use a standard system that many people are familiar with, that is well understood, and for which tools are readily available.

Propositional logic was useful to us but has no chance of ever working beyond specific fixed parameters (such as $n, m = 2, 3$).

Next best thing: (classical) *first-order logic* (FOL)

In my paper with Umberto Grandi (2013) we explored how close we can get to fully modelling *Arrow's Theorem in FOL*.

We also document our—largely unsuccessful—attempts to employ *first-order theorem provers* to get a proof. *But others might do better!*

U. Grandi and U. Endriss. First-Order Logic Formalisation of Impossibility Theorems in Preference Aggregation. *Journal of Philosophical Logic*, 2013.

Initial Observations

- FOL is a natural logic to speak about *binary relations*, such as those used to model preference orders. *Promising!*
- IIA talks about *all* profiles ($=$ complex structures) with certain properties. This has a certain *higher-order feel* to it. *Daunting!*
- FOL cannot express *finiteness*. *Worrisome!*

Language

A key idea is to not talk about profiles (with their internal structure) directly, but to instead introduce the notion of *situation*.

Remark: This corresponds to (and was directly inspired by) the use of numbers to refer to profiles in the SAT approach.

Introduce these *predicate symbols* (with their intuitive meaning):

- $N(z)$: z is an individual
- $A(x)$: x is an alternative
- $S(u)$: u is a situation (referring to a profile)
- $p(z, x, y, u)$: individual z ranks x above y in situation/profile u
- $w(x, y, u)$: society ranks x above y in situation/profile u

Modelling: Social Welfare Functions

We can now write *axioms* (in the logic not the SCT sense of the word!) forcing the intended interpretations, e.g.:

- Individual and collective preferences need to be *linear orders*.
  For instance, p must be interpreted as a *transitive* relation:

$$\forall z.\forall x_1.\forall x_2.\forall x_3.\forall u.\,[\,N(z) \wedge A(x_1) \wedge A(x_2) \wedge A(x_3) \wedge S(u) \;\rightarrow$$
$$(p(z,x_1,x_2,u) \wedge p(z,x_2,x_3,u) \rightarrow p(z,x_1,x_3,u))\,]$$

- The predicates N, A and S must *partition* the domain. That is, any object must belong to exactly one of them:

$$\forall x.[N(x) \vee A(x) \vee S(x)] \;\wedge\; \forall x.[N(x) \rightarrow \neg A(x) \wedge \neg S(x)] \;\wedge\; \cdots$$

Together with a few other simple axioms like this, we can ensure that any model satisfying them must correspond to a SWF (see paper).

The only critical issue is to ensure that models are not too small: we must ensure that the (implicit) *universal domain* assumption holds.

Modelling: Universal Domain Assumption

The universal domain assumption can be modelled, but it's not pretty:

$$
\begin{array}{l}
\forall z.\forall x.\forall y.\forall u.\,[p(z,x,y,u) \rightarrow \exists v.\,[S(v) \wedge p(z,y,x,v) \,\wedge \\
\quad \forall x_1.[p(z,x,x_1,u) \wedge p(z,x_1,y,u) \rightarrow p(z,x_1,x,v) \wedge p(z,y,x_1,v)] \,\wedge \\
\quad \forall x_1.[(p(z,x_1,x,u) \rightarrow p(z,x_1,y,v)) \wedge (p(z,y,x_1,u) \rightarrow p(z,x,x_1,v))] \,\wedge \\
\quad \forall x_1.\forall y_1.[x_1 \neq x \wedge x_1 \neq y \wedge y_1 \neq y \wedge y_1 \neq x \rightarrow \\
\qquad\qquad\qquad\qquad (p(z,x_1,y_1,u) \leftrightarrow p(z,x_1,y_1,v))] \,\wedge \\
\quad \forall z_1.\forall x_1.\forall y_1.\,[z_1 \neq z \rightarrow (p(z_1,x_1,y_1,u) \leftrightarrow p(z_1,x_1,y_1,v))] \quad]]
\end{array}
$$

That is, if there exists a situation u in which individual z ranks x above y, then there must exist a situation v where z ranks y above x and everything else remains the same. Once we ensure the existence of at least one situation, this generates a universal domain.

Modelling: Arrow's Axioms

Modelling Arrow's axioms is fairly easy.

The Pareto condition:

$$S(u) \wedge A(x) \wedge A(y) \rightarrow [\forall z.(N(z) \rightarrow p(z,x,y,u)) \rightarrow w(x,y,u)]$$

Independence of irrelevant alternatives (IIA):

$$\begin{aligned} & S(u_1) \wedge S(u_2) \wedge A(x) \wedge A(y) \;\rightarrow \\ & \qquad [\forall z.(N(z) \rightarrow (p(z,x,y,u_1) \leftrightarrow p(z,x,y,u_2))) \;\rightarrow \\ & \qquad\qquad (w(x,y,u_1) \leftrightarrow w(x,y,u_2))] \end{aligned}$$

Being nondictatorial:

$$\neg\exists z.\, N(z) \wedge \forall u.\forall x.\forall y.\, [S(u) \wedge A(x) \wedge A(y) \wedge p(z,x,y,u) \rightarrow w(x,y,u)]$$

Note: All free variables are understood to be universally quantified.

Modelling: Arrow's Theorem

Let T_{SWF} be the set of axioms determining the theory of SWFs (see paper for full list, including one forcing $m \geqslant 3$). Let T_{ARROW} be the union of T_{SWF} and our three axioms. Then Arrow's Theorem says:

T_{ARROW} *does not have a finite model.*

A shortcoming of this approach is that we cannot reduce this to a statement about some formula being a theorem of FOL. Only if we are willing to fix the number n of individuals, then we can do this (easily).

Thus, for fixed n this approach, in principle, permits a proof of Arrow's Theorem in FOL; and given the availability of complete theorem provers for FOL such a proof can, in principle, be found automatically. However, to date no such proof has been realised in practice.

Second Approach: Modal Logic

Another approach to take is to develop a *new logic* specifically aimed at modelling the aspects of SCT we are interested in.

Modal logic looks like a useful technical framework for doing this.

It is intuitively clear that we can (somehow) devise a modal logic that can capture the Arrovian framework of SWFs, but how to do it exactly is less clear and finding a good way of doing this is a real challenge.

Adopting a semantics-guided approach, we first have to decide:

- what do we take to be our possible worlds?, and
- what accessibility relation(s) should we define?

Exercise: *How would you go about setting up such a modal logic?*

Next, we shall review a specific proposal due to Ågotnes et al. (2011).

T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the Logic of Preference and Judgment Aggregation. *J. Autonomous Agents and Multiagent Systems*, 2011.

Frames

Given: fixed (and finite) N (n individuals) and A (m alternatives)

Each *possible world* consists of

- a profile $\boldsymbol{R}$ and
- an ordered pair (x, y) of alternatives.

There are two *accessibility relations* defined on the possible worlds:

- Two worlds are related via relation $\textsc{prof}$ if their associated pairs are identical (i.e., only their profiles differ, if anything).
- Two worlds are related via relation $\textsc{pair}$ if their associated profiles are identical (i.e., only their pairs differ, if anything).

A *frame* $\langle \mathcal{L}(A)^n \times A^2, \textsc{prof}, \textsc{pair} \rangle$ consists of the set of worlds and the two accessibility relations (all induced by N and A).

Language

The language of the logic has the following *atomic* propositions:

- p_i for every individual $i \in N$
  Intuition: p_i is true at world $\langle \boldsymbol{R}, (x,y) \rangle$ if $x \succ y$ according to R_i
- $q_{(x,y)}$ for every pair of alternatives $(x,y) \in A^2$
  Intuition: $q_{(x',y')}$ is true at world $\langle \boldsymbol{R}, (x,y) \rangle$ if $(x,y) = (x',y')$
- a special proposition σ
  Intuition: σ is true at world $\langle \boldsymbol{R}, (x,y) \rangle$ if society ranks $x \succ y$

The set of *formulas* φ is defined as follows:

$$\varphi \quad ::= \quad p_i \mid q_{(x,y)} \mid \sigma \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\mathrm{PROF}]\varphi \mid [\mathrm{PAIR}]\varphi$$

Disjunction, implication, and diamond-modalities are defined in the usual manner (e.g., $\langle \mathrm{PROF} \rangle \varphi \equiv \neg[\mathrm{PROF}]\neg\varphi$).

Semantics

In modal logic, a *valuation* determines which atomic propositions are true in which world, and a frame and a valuation together define a *model*.

For this logic, the valuation of p_i and $q_{(x,y)}$ is fixed and the valuation of σ will be defined in terms of a SWF F.

So, for given and fixed N and A (and thus for a fixed frame), we now define *truth* of a formula φ at a world $\langle \boldsymbol{R}, (x,y) \rangle$ w.r.t. a SWF F:

- $F, \langle \boldsymbol{R}, (x,y) \rangle \models p_i$ iff $(x,y) \in R_i$
- $F, \langle \boldsymbol{R}, (x,y) \rangle \models q_{(x',y')}$ iff $(x,y) = (x',y')$
- $F, \langle \boldsymbol{R}, (x,y) \rangle \models \sigma$ iff $(x,y) \in F(\boldsymbol{R})$
- $F, \langle \boldsymbol{R}, (x,y) \rangle \models \neg\varphi$ iff $F, \langle \boldsymbol{R}, (x,y) \rangle \not\models \varphi$
- $F, \langle \boldsymbol{R}, (x,y) \rangle \models \varphi \wedge \psi$ iff $F, \langle \boldsymbol{R}, (x,y) \rangle \models \varphi$ and $F, \langle \boldsymbol{R}, (x,y) \rangle \models \psi$
- $F, \langle \boldsymbol{R}, (x,y) \rangle \models [\textsc{prof}]\varphi$ iff $F, \langle \boldsymbol{R'}, (x,y) \rangle \models \varphi$ for all profiles $\boldsymbol{R'}$
- $F, \langle \boldsymbol{R}, (x,y) \rangle \models [\textsc{pair}]\varphi$ iff $F, \langle \boldsymbol{R}, (x',y') \rangle \models \varphi$ for all pairs (x',y')

That is, the operator $[\textsc{prof}]$ is a standard box-modality w.r.t. the relation $\textsc{prof}$ and $[\textsc{pair}]$ is a standard box-modality w.r.t. the relation $\textsc{pair}$.

Decidability

Formula φ is *satisfiable* if there are an F and a world w s.t. $F, w \models \varphi$.

The logic discussed here is *decidable*, i.e., there exists an effective algorithm that will decide whether a given formula is satisfiable:

- First, recall that *the frame is fixed:* to even write down a formula, we need to fix the language, which means fixing N and A.
- Second, observe that the number of possible SWFs is (huge but) *bounded:* there are exactly $m!^{(m!^n)}$ possibilities.
- Third, observe that *model checking is decidable:* there is an effective algorithm for deciding $F, w \models \varphi$ for given F, w, φ.
- Thus, for a given φ we can "just" try model checking for every possible SWF F and every possible world w.

Of course, this is not a practical algorithm. Ågotnes et al. consider complexity questions in more depth and also provide an axiomatisation.

Modelling: The Pareto Condition

We can model the *Pareto condition* as follows:

$$\textsc{pareto} \quad := \quad [\textsc{prof}][\textsc{pair}](p_1 \wedge \cdots \wedge p_n \to \sigma)$$

That is, in every world $\langle \boldsymbol{R}, (x, y)\rangle$ it must be the case that, whenever all individuals rank $x \succ y$ (i.e., all p_i are true), then also society will rank $x \succ y$ (i.e., σ is true).

Write $F \models \varphi$ if $F, w \models \varphi$ for all worlds w.

We have: $F \models \textsc{pareto}$ iff F satisfies the Pareto condition.

Remark: The nesting $[\textsc{prof}][\textsc{pair}]$ amounts to a *universal modality* (you can reach every possible world).

Modelling: Independence of Irrelevant Alternatives

Notation: For any coalition $C \subseteq N$, define p_C as

$$p_C \quad := \quad \bigwedge_{i \in C} p_i \;\wedge \bigwedge_{i \in N \setminus C} \neg p_i.$$

We can now express *IIA:*

$$\textsc{iia} \quad := \quad [\textsc{prof}][\textsc{pair}] \bigwedge_{C \subseteq N} (p_C \wedge \sigma \to [\textsc{prof}](p_C \to \sigma))$$

That is, in every world $\langle \boldsymbol{R}, (x, y) \rangle$ it must be the case that, if exactly the individuals in the group C rank $x \succ y$ (i.e., p_C is true) and society also ranks $x \succ y$ (i.e., σ is true), then for any other profile $\boldsymbol{R'}$ under which still exactly those in C rank $x \succ y$ society also must rank $x \succ y$.

We have $F \models \textsc{iia}$ iff F satisfies IIA.

Modelling: Dictatorships

Finally, we can model what it means for F to be *dictatorial:*

$$\textsc{dictatorial} \quad := \quad \bigvee_{i \in N} [\textsc{prof}][\textsc{pair}](p_i \leftrightarrow \sigma)$$

That is, there exists an individual i (the dictator) such that it is the case that, to whichever world $\langle \boldsymbol{R}, (x,y) \rangle$ we move, society will rank $x \succ y$ (i.e., σ will be true) if and only if i ranks $x \succ y$ (i.e., p_i is true).

We have $F \models \neg\textsc{dictatorial}$ <u>iff</u> F is nondictatorial.

Modelling: Arrow's Theorem

Write $\models \varphi$ if $F \models \varphi$ for all SWFs F (for the fixed sets N and A).

We can now state *Arrow's Theorem:*

$$\text{If } |A| \geqslant 3, \text{ then } \models \neg(\textsc{pareto} \wedge \textsc{iia} \wedge \neg\textsc{dictatorial}).$$

Note that this does *not* mean that we have a proof within this logic, although the completeness result of Ågotnes et al. regarding their axiomatisation means that such a proof is feasible in principle.

In my paper with Giovanni Ciná (2016), we've been able to sketch such a (Hilbert-style) *proof* of Arrow's Theorem for SCFs in a similar logic.

Remark: Importantly, the above is a statement of Arrow's Theorem only for *fixed* (but arbitrary) choices of *N and A*.

G. Ciná and U. Endriss. Proving Classical Theorems of Social Choice Theory in Modal Logic. *J. Autonomous Agents and Multiagent Systems*, 2016.

Higher-Order Logic Proof Assistants

There also has been work on verifying the correctness of known proofs of results in SCT using HOL proof assistants such as *Isabelle* or *Coq*.

Nipkow's paper on Arrow's Theorem and G-S is an example.

T. Nipkow. Social Choice Theory in HOL. *Journal of Automated Reasoning*, 2009.

Formal Verification

A further logic-based application is the use of *model checking* to verify the correctness of *implementations* (e.g., in Java) of voting rules.

Beckert et al. (2017) give an introduction to this topic.

B. Beckert, T. Bormer, R. Goré, M. Kirsten, and C. Schürmann. An Introduction to Voting Rule Verification. In *Trends in COMSOC*. AI Access, 2017.

Summary

We've seen different approaches to *modelling* features of SCT *in logic*, providing different degrees of support for *automated reasoning:*

- propositional logic (for small sets of individuals/alternatives)
- first-order logic (for arbitrary numbers of individuals/alternatives)
- modal logic (specifically designed for this job)

We are left with (at least) these questions and challenges:

- don't fix the *set of individuals* (and alternatives) in the language
- model the *universal domain* assumption in an elegant manner
- better support *automated reasoning* for the richer languages

What next? Social choice in richer models of decision making.