Automated Reasoning in Social Choice Theory: Arrow's Theorem

Umberto Grandi Ulle Endriss

February 18, 2009

Outline

1. Arrow's Theorem: an overview and a useful lemma;
2. First-order axiomatisation for automated theorem proving:
   - express Arrow's hypothesis;
   - problems dealing with infinite models;
   - under certain conditions Arrow's Theorem can be proved.
3. Prover9: results with an automated theorem prover.

Social Welfare Functions

- I a set of individuals, A a set of alternatives;
- P_i is a linear order over alternatives A;
- $\mathcal{L}(A)$ is the set of all linear orders on A;

Definition

A social welfare function (SWF) for A and I is a function $w : \mathcal{L}(A)^I \longrightarrow \mathcal{L}(A)$

w associate to every preference profile $\underline{P} = (P_1, \ldots, P_n)$ a "social order" $w(\underline{P})$

Arrow's conditions:

- Unanimity (**UN**): if aP_ib for every $i \in I$ then $aw(\underline{P})b$;
- Independence of Irrelevant Alternatives (**IIA**): given two preference profiles $\underline{P}$ and $\underline{Q}$, if aP_ib if and only if aQ_ib for every $i \in I$, then $aw(\underline{P})b$ if and only if $aw(\underline{Q})b$;
- Non-dictatorship (**NDIC**): there is no individual i such that for every profile $\underline{P}$ the social order $w(\underline{P}) = P_i$.

Arrow's Theorem

Theorem (Arrow, 1950)

If A and I are finite and non-empty, and $|A| \geqslant 3$, then there is no social welfare function for I and A that satisfies **UN**, **IIA** *and* **NDIC**.

In a recent paper by Lin and Tang (2008) present an inductive proof:

If there exist a SWF for $|A| = m + 1$ and $|I| = n$ satisfying Arrow's conditions then there exist a SWF for $|A| = m$ and $|I| = n$ satisfying the same properties.

$$\Downarrow$$

If Arrow's Theorem holds for $|A| = 3$ and $|I| = n$ then it holds for $|A| = m$ and $|I| = n$ for every m

K. Arrow, A Difficulty in the Concept of Social Welfare. The Journal of Political Economy, 1950.

F. Lin and P. Tang. Computer-Aided Proofs of Arrow's and other Impossibility Theorems. AAAI 2008.

Infinite Number of Alternatives

Lemma

If there exist a SWF for $|A| \geqslant 3$ and I that satisfy **UN**, **IIA** *and* **NDIC** *then there exist a SWF for $|A'| = 3$ and I that satisfies the same properties.*

Proof.

Let A' be a subset of A containing 3 different elements, and $\underline{P}$ be a profile on A'
Define $w' : \mathcal{L}(A')^I \to \mathcal{L}(A')$ as:

$$a\, w'(\underline{P})\, b \;:\Leftrightarrow\; a\, w(\underline{P^e})\, b$$

where $\underline{P^e}$ is any extension of $\underline{P}$ to a linear order on A.

- it does not depend on the chosen extension, due to **IIA** of w;
- w' satisfies **UN** and **IIA**;
- Suppose i is a dictator for w', we show that it is also a dictator for w.
  Take $x, y \in A$ and xP_iy in $\underline{P}$. Be a_1, a_2 in A' different than x and y, define $\underline{P'}$ from $\underline{P}$ by moving xP_ja_1 and a_2P_jy for every j.
  By **UN** we get $xw(\underline{P'})a_1$ and $a_2w(\underline{P'})y$ and by "local" dictatorship $a_1w(\underline{P'})a_2$. By **IIA** and transitivity we then get $xw(\underline{P})y$.

First-Order Logic

Now we present a formal system to model Arrow's theorem:

First-order logic because:

- Natural language to talk about linear orders;
- First-order automated theorem provers are more developed.

PROBLEM: second-order quantification?

UN: $\forall$ preference profiles $\underline{P}$ $\forall$ alternatives x, y ($\forall$ individual i xP_iy)$\rightarrow$ $(xw(\underline{P})y)$

SOLUTION:

A situation is a "name" for a preference profile:

$$s \rightarrow \underline{P}^s$$

where s is an element of the model and $\underline{P}^s$ the profile associated.

UN: ($\forall$ situation s $\forall$ alternatives x, y ($\forall$ individual i xP_i^sy)$\rightarrow$ $(xw(\underline{P}^s)y)$)

now is almost a first order statement.

Language

To express statements of this kind we need:

- unary relations to mark individuals $I(z)$, alternatives $A(x)$ and situations $S(u)$;
- constant symbols a_1, a_2, a_3 for 3 alternatives, i_1 and s_1 for an individual and a situation;
- a relation $p(z, x, y, u)$ to represent the linear order P_z^u of z in situation u (arity 4);
- a relation $w(x, y, u)$ to represent the social outcome $w(\underline{P}^u)$ for every situation u.

$$\mathcal{L} = \{a_1,\ a_2,\ a_3,\ i_1,\ s_1,\ I^{(1)},\ A^{(1)},\ S^{(1)},\ w^{(3)},\ p^{(4)}\}$$

Axioms:

LIN$_p$: p is a linear order for every individual in every situation

- $I(z) \wedge S(u) \wedge A(x) \wedge A(y) \rightarrow (p(z, x, y, u) \vee p(z, y, x, u) \vee x = y)$
- $I(z) \wedge S(u) \wedge A(x) \rightarrow \neg p(z, x, x, u)$
- $I(z) \wedge S(u) \wedge A(x_1) \wedge A(x_2) \wedge A(x_3) \wedge p(z, x_1, x_2, u) \wedge p(z, x_2, x_3, u) \rightarrow p(z, x_1, x_3, u)$
- $p(z, x, y, u) \rightarrow (I(z) \wedge A(x) \wedge A(y) \wedge S(u))$

Axioms I

- **LIN**$_w$: w is a linear order in every situation
  - $S(u) \wedge A(x) \wedge A(y) \rightarrow (w(x,y,u) \vee w(y,x,u)) \vee x = y$
  - $S(u) \wedge A(x) \rightarrow \neg w(x,x,u)$
  - $S(u) \wedge A(x_1) \wedge A(x_2) \wedge A(x_3) \wedge w(x_1,x_2,u) \wedge w(x_2,x_3,u) \rightarrow w(x_1,x_3,u)$
  - $w(x,y,u) \rightarrow (A(x) \wedge A(y) \wedge S(u))$
- **MIN**: A and I are non-empty and there are at least 3 alternatives
  - $A(a_1) \wedge A(a_2) \wedge A(a_3) \wedge I(i_1) \wedge S(s_1)$
  - $\neg(a_1 = a_2) \wedge \neg(a_1 = a_3) \wedge \neg(a_2 = a_3)$
- **PART**: I, A and S form a partition
  - $A(x) \rightarrow (\neg I(x) \wedge \neg S(x))$
  - $I(x) \rightarrow (\neg A(x) \wedge \neg S(x))$
  - $S(x) \rightarrow (\neg I(x) \wedge \neg A(x))$
  - $A(x) \vee I(x) \vee S(x)$
- **INJ**: two different situations encode different orders
  - $S(u) \wedge S(v) \wedge (u \neq v) \rightarrow \exists z,\, x,\, y\ \ [I(z) \wedge A(x) \wedge A(y) \wedge p(z,x,y,u) \wedge p(z,y,x,v)]$

Axioms II: Permutations

A hidden hypothesis of Arrow's Theorem is universal domain:
a SWF is defined on every possible preference profile in $\mathcal{L}(A)^I$.

Even models with only a single situation are allowed by our axioms.
We have to rule out such "small" models with the next **PERM** axiom:

- $p(z,x,y,u) \rightarrow \exists v \{S(v) \wedge p(z,y,x,v) \wedge$
  $\forall x_1[p(z,x,x_1,u) \wedge p(z,x_1,y,u) \rightarrow p(z,x_1,x,v) \wedge p(z,y,x_1,v)] \wedge$
  $\forall x_1[(p(z,x_1,x,u) \rightarrow p(z,x_1,x,v)) \wedge (p(z,y,x_1,u) \rightarrow p(z,y,x_1,v))] \wedge$
  $\forall x_1 \forall y_1[(x_1 \neq x) \wedge (y_1 \neq y) \wedge p(z,x_1,y_1,u) \rightarrow p(z,x_1,y_1,v)]$
  $\forall z_1,\ x,\ y\ [(z_1 \neq z) \wedge I(z_1) \wedge A(x) \wedge A(y) \rightarrow (p(z_1,x,y,u) \leftrightarrow p(z_1,x,y,v))]\}$

If in situation u the order of individual z is:

$$\ldots x \succ_z^u a \succ_z^u b \succ_z^u y \ldots$$

then there exist a situation v where these two alternatives are swapped:

$$\ldots y \succ_z^v a \succ_z^v b \succ_z^v x \ldots$$

Axioms III: Arrow's Conditions

Call T_{SWF} the axioms presented so far.

Add the following axioms and call the resulting theory T_{ARROW}:

- **UN**: $S(u) \wedge A(x) \wedge A(y) \rightarrow [(\forall z \;\; I(z) \rightarrow p(z,x,y,u)) \rightarrow w(x,y,u)]$
- **IIA**: $S(u_1) \wedge S(u_2) \wedge A(x) \wedge A(y) \rightarrow$
  $[(\forall z \;\; I(z) \rightarrow (p(z,x,y,u_1) \leftrightarrow p(z,x,y,u_2))) \rightarrow (w(x,y,u_1) \leftrightarrow w(x,y,u_2))]$
- **NDIC**: $I(z) \rightarrow [\exists x,\, y,\, u \;\; A(x) \wedge A(y) \wedge (x \neq y) \wedge S(u) \wedge p(z,x,y,u) \wedge w(y,x,u)]$

Arrow's Theorem can be restated as:

Theorem

T_{ARROW} has no finite models.

Models of Social Welfare Functions

To every SWF w for $|A| \geqslant 3$ and I we can associate a model $\mathcal{M}_w$ of T_{SWF}

$$\mathcal{M}_w = (M, a_1, a_2, a_3, i_1, s_1, I, A, S, p, w)$$

- the universe $M = A \cup I \cup \mathcal{L}(A)^I$, corresponding to the three unary relations A, I and S;
- a_1, a_2, a_3 are three different elements of A, i_1 in I and s_1 in $\mathcal{L}(A)^I$;
- $(z, x, y, u) \in p \Leftrightarrow x \, P_z^u \, y$;
- $(x, y, u) \in w \Leftrightarrow x \, w(\underline{P}^u) \, y$.

It is easy to see that $\mathcal{M}_w$ is a model of T_{SWF}

in particular the **PERM** axiom is valid
because the model is "built" on a universal domain.

Completeness

If A is infinite, $\mathcal{M}_w$ is not the unique model we can build from w:

Definition

The set $\mathcal{L}(A)$ can be identified with the set $S(A)$ of all permutations over A.
A transposition is a permutation that exchange only two elements.
A subset $G \subseteq S(A)$ is closed under transpositions if when $g \in G$ then $g \circ \tau \in G$ for every transposition τ.

Remark: if A is finite, $S(A)$ is the only subset closed under transpositions.

$$\begin{array}{c} \text{For every choice of a } G_i \\ \text{c.u.t for every individual } i \end{array} \quad \Rightarrow \quad \begin{array}{c} \text{model } \mathcal{M}_w \text{ with domain} \\ A \cup I \cup \prod_{i \in I} G_i \end{array}$$

The **PERM** axiom remains true.

Proposition (Completeness)

$\mathcal{M}$ is a model of T_{SWF} if and only if there exists two non empty sets A and I, with $|A| \geqslant 3$, and a SWF w for A and I such that $\mathcal{M} = \mathcal{M}_w$.

Infinite Number of Individuals

If I is infinite then there exists a SWF for I and $|A| \geqslant 3$ that satisfies **UN**, **IIA** and **NDIC** (Fishburn, 1970)

$$\Downarrow$$

T_{ARROW} is consistent.

1. In the same paper Fishburn proves that if Arrow's conditions hold then the number of individuals must be infinite: not directly expressible in first-order logic.
2. Fix the number of individuals (new constants $i_1, \ldots, i_n$) and add the axioms:
   - $i_k \neq i_j$ for every $k \neq j$;
   - $I(i_1) \wedge \cdots \wedge I(i_n)$;
   - $I(z) \rightarrow (z = i_1) \vee \cdots \vee (z = i_n)$.

   Obtaining the theory T^n_{SWF}.

P. Fishburn, Arrow's Theorem: Concise Proof and Infinite Voters. Journal of Economic Theory, 1970.

Fixed Number of Individuals

Same completeness result:

Proposition

$\mathcal{M}$ is a model of T^n_{SWF} if and only if there exists two non empty sets A and I, with $|A| \geqslant 3$ and $|I| = n$, and a SWF w for A and I such that $\mathcal{M} = \mathcal{M}_w$.

And now there is no escape from Arrow's Theorem:

Proposition

If w is a SWF for A and I with $|A| \geqslant 3$ and $|I| = n$ then:

$$\mathcal{M}_w \models \neg(\mathbf{UN} \wedge \mathbf{IIA} \wedge \mathbf{NDIC}).$$

Therefore for every n:

$$T^n_{SWF} \vdash \neg(\mathbf{UN} \wedge \mathbf{IIA} \wedge \mathbf{NDIC})$$

Possibly different proofs

Prover9

To summarize, an automated theorem prover can:

- prove Arrow's theorem for a fixed number of individuals and alternatives;
- prove Arrow's theorem fixing only the number of individuals;
- prove Fishburn's Theorem?

We implemented our axiomatisation using Prover 9 (successor of Otter) with the following results:

- even the easiest case of 3 alternatives and 2 individuals exceeds the search space limits;
- based on some pre-existent formalisations we designed a guided step-by-step proof: exceeds the search space limits;
- we were able to understand why.

```
36 -p(x,y,z,u) | A(y).  [clausify(8)].
37 -p(x,y,z,u) | A(z).  [clausify(8)].
38 -p(x,y,z,u) | S(u).  [clausify(8)].
39 -w(x,y,z) | A(x).  [clausify(9)].
40 -w(x,y,z) | A(y).  [clausify(9)].
41 -w(x,y,z) | S(z).  [clausify(9)].
42 -I(x) | b1 = x | b2 = x.  [clausify(10)].
43 -p(x,y,z,u) | S(f(y,z,x,u)).  [assumption].
44 -p(x,y,z,u) | p(x,z,y,f(y,z,x,u)).  [assumption].
45 -p(x,y,z,u) | -p(x,y,w,u) | -p(x,w,z,u) | p(x,w,y,f(y,z,x,u)).  [assumption].
46 -p(x,y,z,u) | -p(x,y,w,u) | -p(x,w,z,u) | p(x,z,w,f(y,z,x,u)).  [assumption].
47 -p(x,y,z,u) | -p(x,w,y,u) | p(x,w,y,f(y,z,x,u)).  [assumption].
48 -p(x,y,z,u) | -p(x,z,w,u) | p(x,z,w,f(y,z,x,u)).  [assumption].
49 -p(x,y,z,u) | w = y | v5 = z | -p(x,w,v5,u) | p(x,w,v5,f(y,z,x,u)).  [assumption].
50 -p(x,y,z,u) | -I(w) | -A(v5) | -A(v6) | w = x | -p(w,v5,v6,u) | p(w,v5,v6,f(y,z,x,u)).  [assumption].
51 -p(x,y,z,u) | -I(w) | -A(v5) | -A(v6) | w = x | p(w,v5,v6,u) | -p(w,v5,v6,f(y,z,x,u)).  [assumption].
52 -S(x) | -S(y) | y = x | I(f1(x,y)).  [clausify(11)].
53 -S(x) | -S(y) | y = x | A(f2(x,y)).  [clausify(11)].
```

All proofs can be divided in two steps:

- there exists a preference profile satisfying certain conditions — NO
- using Arrow's axioms derive a contradiction — OK

To prove the first statement the prover has to guess the right swapping sequence, for k swaps there are $k \cdot |I| \cdot \binom{|A|}{2}$ possible sequences, and draw the right conclusions at every step.

A possible solution is a guided use of the **PERM** axiom (using hints).

Related Works

- Lin and Tang (2008) used a SAT-solver to check the base case of 3 alternatives and 2 individuals, using two inductive lemmas to obtain the full statement of Arrow's Theorem in the case of finite number of alternatives and individuals;
- Agotnes *et al.* (2008) introduced a modal logic tailored for the more general framework of judgment aggregation, and showed that Arrow's Theorem can be proved in their logic. Both the number of individuals and the number of alternatives are fixed in the language, and no automated proving procedure for this logic.
- Another approach has been followed by Nipkov and Wiedijk (2008, 2007) using higher-order automated theorem checker to formalize different proofs of Arrow's Theorem in the finite case.

F. Lin and P. Tang. Computer-Aided Proofs of Arrow's and other Impossibility Theorems. AAAI 2008.

Agotnes, van der Hoek, Woolridge, Reasoning About Judgment and Preference Aggregation. AAMAS 2007.

T. Nipkov, Social Choice Theory in HOL. JAR, 2008.

F. Wiedijk, Arrow's Impossibility Theorem. Formalized Mathematics, 2007.

Conclusion and Future Works

In this talk:

- We presented a first-order axiomatisation of SWF that allow to express Arrow's conditions.
- A completeness result with respect to SWF has been proved
- We dealt with the case of an infinite number of alternatives, and showed that Arrow's Theorem is provable in the case of a fixed number of individuals.
- We discussed the results of our implementation with Prover 9.

Other results and future works:

- We were able to check the axiomatization using Mace4 and Clausetester.
- Mace4 can generate models containing all possible orders (only one individual).
- Formalize other impossibility results (Gibart-Satterwhite's Theorem, Sen's Liberal Paradox...).
- Formalize known possibility results: Black's Theorem, Sen's value restriction.
- Test unknown possibility results using Prover9 and Mace4 on a weaker version of our axioms.