First-Order Logic Formalisation of Arrow's Theorem

Umberto Grandi   Ulle Endriss

July 1, 2009

Outline

1. Arrow's Theorem: classical formulation and a useful lemma;
2. First-Order Axiomatisation:
    - Express Arrow's conditions;
    - Completeness result for social welfare functions;
    - Problems dealing with infinite models.
3. Automated Reasoning: state under which conditions Arrow's Theorem can be proved automatically and report on some preliminary results with an automated theorem prover.

Social Welfare Functions

- I a set of individuals, A a set of alternatives;
- $P_i \in \mathcal{L}(A)$ is a linear order over alternatives 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" $w(\underline{P})$

Social Welfare Functions

- I a set of individuals, A a set of alternatives;
- $P_i \in \mathcal{L}(A)$ is a linear order over alternatives 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 individual i then $aw(\underline{P})b$;
- Independence of Irrelevant Alternatives (**IIA**): the relative social ranking of two alternatives a and b depends only on the relative ranking of these two alternatives by the individuals;
- 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**.

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.

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 exists 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.

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 exists 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

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.*

First-Order Logic

First-order logic is a natural language to talk about orders and first-order automated theorem provers are more developed than for other systems.

PROBLEM: Second-order quantification?

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

First-Order Logic

First-order logic is a natural language to talk about orders and first-order automated theorem provers are more developed than for other systems.

PROBLEM: Second-order quantification?

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

SOLUTION: Introduce a set of situations as "names" for preference profiles:

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

First-Order Logic

First-order logic is a natural language to talk about orders and first-order automated theorem provers are more developed than for other systems.

PROBLEM: Second-order quantification?

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

SOLUTION: Introduce a set of situations as "names" for preference profiles:

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

We define the theory of social welfare functions T_{SWF} using the following language:

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

And axioms like the followings:

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

Axioms II: Permutation

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

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

Axioms II: Permutation

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

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

If in situation u the order of individual z is:

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

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

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

Axioms III: Arrow's Conditions

Add to T_{SWF} 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**: $I(z) \rightarrow [\exists x,\, y,\, u \;\; A(x) \wedge A(y) \wedge (x \neq y) \wedge S(u) \wedge p(z,x,y,u) \wedge w(y,x,u)]$

Axioms III: Arrow's Conditions

Add to T_{SWF} 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}

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

With universe $M = A \cup I \cup \mathcal{L}(A)^I$, corresponding to the three unary relations A, I, S.

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}

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

With universe $M = A \cup I \cup \mathcal{L}(A)^I$, corresponding to the three unary relations A, I, S.

If A is infinite, for every choice of $G_i \subseteq \mathcal{L}(A)$ closed under transpositions we can use $S = \prod_i G_i$ in the definition above to build different models of $\mathcal{M}_w$.

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}

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

With universe $M = A \cup I \cup \mathcal{L}(A)^I$, corresponding to the three unary relations A, I, S.

If A is infinite, for every choice of $G_i \subseteq \mathcal{L}(A)$ closed under transpositions we can use $S = \prod_i G_i$ in the definition above to build different models of $\mathcal{M}_w$.

Proposition (Completeness)

$\mathcal{M}$ is a model of T_{SWF} if and only if there exist 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 I

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.

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

Infinite Number of Individuals I

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. A first way out is to fix the number of individuals adding new constants and axioms. For every n, this theory T^n_{SWF} can derive an instance of Arrow's Theorem:

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

Infinite Number of Individuals I

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. A first way out is to fix the number of individuals adding new constants and axioms. For every n, this theory T^n_{SWF} can derive an instance of 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 $T^n_{\text{SWF}} \vdash \neg(\mathbf{UN} \wedge \mathbf{IIA} \wedge \mathbf{NDIC})$

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

Infinite Number of Individuals II

1. A second approach is to formalize Kirman and Sondermann's Theorem:

Theorem

If a SWF satisfies **UN** *and* **IIA** *then the set of winning coalitions is an ultrafilter over I.*

We give axioms for this statement, that holds true on every model $\mathcal{M}_w$.
Therefore it is provable by T_{SWF}.

Infinite Number of Individuals II

1. A second approach is to formalize Kirman and Sondermann's Theorem:

Theorem

If a SWF satisfies **UN** *and* **IIA** *then the set of winning coalitions is an ultrafilter over I.*

We give axioms for this statement, that holds true on every model $\mathcal{M}_w$.
Therefore it is provable by T_{SWF}.

To summarize, an automated theorem prover can:

- prove Arrow's theorem fixing only the number of individuals: possibly different proofs!
- prove the more general result by Kirman and Sondermann, that implies Arrow's Theorem in full generality.

Conclusion and Future Work

We implemented our axiomatisation using Prover 9 (successor of Otter) and E-prover:

- We were able to prove very simple results on instantiated domains;
- Even the simplest case of 2 individuals and 3 alternatives of Arrow's Theorem exceeded the search space limit.

Conclusion and Future Work

We implemented our axiomatisation using Prover 9 (successor of Otter) and E-prover:

- We were able to prove very simple results on instantiated domains;
- Even the simplest case of 2 individuals and 3 alternatives of Arrow's Theorem exceeded the search space limit.

In this work:

- We presented a first-order axiomatisation of SWF that allow to express Arrow's conditions and proved a completeness result;
- We dealt with the case of an infinite number of alternatives, and stated under which conditions Arrow's Theorem is provable in the case of an infinite number of individuals.

Conclusion and Future Work

We implemented our axiomatisation using Prover 9 (successor of Otter) and E-prover:

- We were able to prove very simple results on instantiated domains;
- Even the simplest case of 2 individuals and 3 alternatives of Arrow's Theorem exceeded the search space limit.

In this work:

- We presented a first-order axiomatisation of SWF that allow to express Arrow's conditions and proved a completeness result;
- We dealt with the case of an infinite number of alternatives, and stated under which conditions Arrow's Theorem is provable in the case of an infinite number of individuals.

For the future:

- Formalise of other impossibility and possibility results (Gibbard-Satterthwaite's Theorem, Sen's Liberal Paradox...).
- Explore new results automatically testing weaker versions of our axioms.

Axioms

Axioms

- **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
  - $S(u) \wedge S(v) \wedge (u \neq v) \rightarrow \exists z,\ x,\ y\ \ [I(z) \wedge A(x) \wedge A(y) \wedge p(z,x,y,u) \wedge p(z,y,x,v)]$