Overview of Automated Reasoning Techniques in Social Choice Theory

Umberto Grandi

Institut de Recherche en Informatique de Toulouse (IRIT)
University of Toulouse 1 Capitole

28 March 2019

Overview

1. Basic definitions of social choice theory
2. Main motivation: proof assistance and theorem discovery
3. Propositional logic and SAT solvers
4. Modal logic and higher-order logics
5. First-order logic
6. Possible intersections with AGAPE

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 $w : \mathcal{L}(A)^I \to \mathcal{L}(A)$

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

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 $w : \mathcal{L}(A)^I \to \mathcal{L}(A)$

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

Arrow's conditions:

- **Unanimity (UN)**: if aP_ib for every individual i then $aw(\boldsymbol{P})b$;
- **Independence of Irrelevant Alternatives (IIA)**: the relative social ranking of two alternatives a and b depends only on the relative ranking of a and b by the individuals;
- **Non-dictatorship (NDIC)**: there is no individual i such that for every profile $\boldsymbol{P}$ the social order $w(\boldsymbol{P}) = P_i$.

Two famous theorems in social choice

Arrow's Theorem (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**.

The original proof contained a mistake! Pointed out by Blau (1957) and fixed in the second edition of the book. This motivated the use of automated proof-checkers based on higher-order logic (say a lighter version of set theory) like Isabelle (Nipkow, 2009) and Mizar (Wiedijk, 2007).

Two famous theorems in social choice

Arrow's Theorem (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**.

The original proof contained a mistake! Pointed out by Blau (1957) and fixed in the second edition of the book. This motivated the use of automated proof-checkers based on higher-order logic (say a lighter version of set theory) like Isabelle (Nipkow, 2009) and Mizar (Wiedijk, 2007).

Let now a voting rule be a function $r : \mathcal{L}(A)^I \to A$ associating a winning alternative to every profile of linear orders:

> **Strategy-proofness**: There is no profile $\boldsymbol{P}$, voter i, and linear order P_i' such that $r(\boldsymbol{P}_{-i}, P_i')\ P_i\ r(\boldsymbol{P})$.

Gibbard-Sattertwaite Theorem (1973-1975)

If A and I are finite and non-empty, and $|A| \geqslant 3$, then any voting rule for I and A that is onto and strategy-proof is a dictatorship.

Proof assistance and theorem discovery

Most proofs in social choice theory are combinatorial: voters and alternatives are finite discrete sets, no probability is involved in classical setting, properties and axioms are almost expressed in relational language. Observations:

1. Theorems and proofs often hinges on specific modelling hypothesis: should we consider linear orders, weak orders, partial orders? A social welfare function or social choice rule? Potentially many declinations of Arrow's theorem, each with a different proof. This is good for paper writing, but can the process be automated?
2. Some combinatorial proofs are very hard and resisted researchers for a long time. Computer aided proofs could prove useful.
3. On a completely different perspective, what is the logic of social welfare functions (social choice rules)? Can we come up with a formalism and some axioms so that Arrow's and similar theorems can be derived in these logics? Teaser: independence and strategy proofness have an intriguing universal quantifier on profiles (linear orders)...

Propositional Logic and SAT solvers

Tang and Lin (2009) proposed an inductive proof of Arrow's theorem:

If there exists 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

Propositional Logic and SAT solvers

Tang and Lin (2009) proposed an inductive proof of Arrow's theorem:

If there exists 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

Followed by a model-check of the base case of 3 alternatives and 2 individuals feeding a SAT solver with instantiations of formulas like the Pareto axiom:

$$\text{FORALL } a, b, s[\text{FORALL } x \;\; p(x, a, b, s)] \rightarrow w(a, b, s)$$

Some numbers: there are $\sim 10^{28}$ social welfare functions for the base case, the SAT solver gave an answer in < 1 second for 35k variables and 100k clauses.

Tang and Lin. Computer-Aided Proofs of Arrow's and other Impossibility Theorems. AIJ, 2009.

Theorem discovery via SAT solvers and more

- Geist and Endriss (2011), in the setting of ranking sets of objects, proved an inductive step for a class of formulas in a logical language that encodes classical axioms, and run SAT on the base case of all combinations of axioms, finding 84 axiom-minimal impossibilities.

Geist and Endriss. Automated Search for Impossibility Theorems in Social Choice Theory: Ranking Sets of Objects. *JAIR*, 2011. IJCAI-JAIR Best Paper Prize 2016

Geist and Peters. Computer-Aided Methods for Social Choice Theory. In *Trends in Computational Social Choice*, U. Endriss (ed), 2017.

Theorem discovery via SAT solvers and more

- Geist and Endriss (2011), in the setting of ranking sets of objects, proved an inductive step for a class of formulas in a logical language that encodes classical axioms, and run SAT on the base case of all combinations of axioms, finding 84 axiom-minimal impossibilities.
- Felix Brandt's research group in Munich took over, several papers regularly published typically with inductive proofs + MUS extraction to produce a human-readable proof. Close up on how they model strategy-proofness:

$$\varphi_{\text{strategyproof}} \equiv \bigwedge_{i \in N} \bigwedge_{R \in \mathcal{R}} \bigwedge_{\substack{R' \in \mathcal{R} \\ R'(j) = R(j) \\ \forall j \neq i}} \bigwedge_{\substack{x, y \in A \\ x \succ_i y}} (v_{R,y} \rightarrow \neg v_{R',x}).$$

Geist and Endriss. Automated Search for Impossibility Theorems in Social Choice Theory: Ranking Sets of Objects. *JAIR*, 2011. IJCAI-JAIR Best Paper Prize 2016

Geist and Peters. Computer-Aided Methods for Social Choice Theory. In *Trends in Computational Social Choice*, U. Endriss (ed), 2017.

Theorem discovery via SAT solvers and more

- Geist and Endriss (2011), in the setting of ranking sets of objects, proved an inductive step for a class of formulas in a logical language that encodes classical axioms, and run SAT on the base case of all combinations of axioms, finding 84 axiom-minimal impossibilities.
- Felix Brandt's research group in Munich took over, several papers regularly published typically with inductive proofs + MUS extraction to produce a human-readable proof. Close up on how they model strategy-proofness:

$$\varphi_{\text{strategyproof}} \equiv \bigwedge_{i \in N} \bigwedge_{R \in \mathcal{R}} \bigwedge_{\substack{R' \in \mathcal{R} \\ R'(j)=R(j) \\ \forall j \neq i}} \bigwedge_{\substack{x,y \in A \\ x \succ_i y}} (v_{R,y} \rightarrow \neg v_{R',x}).$$

- Interesting literature for AGAPE: Frechette et al. (2016) encoded the reverse spectrum auction with SAT solvers, Caminati et al. (2015) verified combinatorial Vickrey auctions with HOL provers.

Geist and Endriss. Automated Search for Impossibility Theorems in Social Choice Theory: Ranking Sets of Objects. *JAIR*, 2011. IJCAI-JAIR Best Paper Prize 2016

Geist and Peters. Computer-Aided Methods for Social Choice Theory. In *Trends in Computational Social Choice*, U. Endriss (ed), 2017.

Modal logic

Troquard et al. (2011) (follow-up by Ciná and Endriss), consider reported profiles of preferences as possible worlds of a Kripke model:

- atomic variables $p^i_{x>y}$ controlled by each agent i
- atoms $x \in A$ to specify the winning alternative (encoding the voting rule)
- $\Diamond_C \varphi$ stands for coalition C has a strategy such that φ holds provided all players outside C stick to their current strategies
- $\blacklozenge_i \varphi$ stands for i prefers a profile where φ is true to the current one

Modal logic

Troquard et al. (2011) (follow-up by Ciná and Endriss), consider reported profiles of preferences as possible worlds of a Kripke model:

- atomic variables $p^i_{x>y}$ controlled by each agent i
- atoms $x \in A$ to specify the winning alternative (encoding the voting rule)
- $\Diamond_C \varphi$ stands for coalition C has a strategy such that φ holds provided all players outside C stick to their current strategies
- $\blacklozenge_i \varphi$ stands for i prefers a profile where φ is true to the current one

Strategy-proofness of a voting rule r looks like this:

$$\rho(r) \rightarrow \bigwedge_{\text{all profiles } < \text{ over A}} [\text{true}(<) \rightarrow \texttt{DOM})]$$

Where $\rho(r)$ states that the model correctly encodes r, true$(<)$ "reifies" the sincere preference profile $<$ of the agents, and

$$\texttt{DOM} = \bigwedge_i \Box_{N \setminus i} \bigvee_{x \in A} (x \wedge \Box_i \blacklozenge_i x)$$

How to read this: "for any agent i, for any deviation of all other agents, there exists an alternative x that is the winner at the reported profile, and for all unilateral deviations of i, i prefers x.

Modal logic - Discussion

Pros:

- Strategy-proofness for voting rules can be equivalent to other axiomatic properties (more formalisation-friendly)!
- Off-the-shelves techniques in epistemic logic might be easier to plug. After-all, strategy-proofness depends on the information available to agents

Cons:

- The number of voters and alternatives has to be specified in the language (this is almost unavoidable, and probably not a big issue?)
- Universal quantifications on profiles are exogenously coded with big conjunctions (n players and m alternatives means $(m!)^n$ profiles!)
- Logics are "ad-hoc", dedicated solvers typically do not exist

Troquard, van der Hoek, and Wooldridge. Reasoning about social choice functions. *Journal of Philosophical Logic*, 2011.
Agotnes, van der Hoek, and Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multiagent Systems*, 2011

Ciná and Endriss. Proving Classical Theorems of Social Choice Theory in Modal Logic.

Autonomous Agents and Multiagent Systems, 2016.

Dynamic Logic of Propositional Assignments

An interesting formalism: it can be used to write complex axioms easily, which are then transformed into (long) propositional formulas ready for a SAT solver.

- Variables $p^i_{x>y}, p^r_{x>y}$ for individual preferences and outcomes
- Constraints enforcing linear, weak, or partial orders...
- Basic program flipping variables, complex programs are able to encode most existing voting rules, and verify their axiomatic properties

$$\mathsf{slater}_{\mathrm{IC}}(\mathbb{B}) := \mathsf{maj}(\mathbb{B})\,;\bigcup_{0\leq d\leq m}\left(\mathsf{H}(\mathrm{IC},\mathbb{O},\geq d)?\,;\mathsf{flip}^1(\mathbb{O})^d\right);\mathrm{IC}?.$$

Dynamic Logic of Propositional Assignments

An interesting formalism: it can be used to write complex axioms easily, which are then transformed into (long) propositional formulas ready for a SAT solver.

- Variables $p^i_{x>y}, p^r_{x>y}$ for individual preferences and outcomes
- Constraints enforcing linear, weak, or partial orders...
- Basic program flipping variables, complex programs are able to encode most existing voting rules, and verify their axiomatic properties

$$\mathsf{slater}_{\mathrm{IC}}(\mathbb{B}) := \mathsf{maj}(\mathbb{B})\,;\bigcup_{0\leq d\leq m}\left(\mathsf{H}(\mathrm{IC},\mathbb{O},\geq d)?\,;\mathsf{flip}^1(\mathbb{O})^d\right);\mathrm{IC}?.$$

Main problems:

- Automated translation sofware not existing yet
- Need to specify the number of players and alternatives in the language
- Impossibility theorems not provable (voting rules are programs, and quantification over programs is not possible)

Arianna Novaro, Umberto Grandi and Andreas Herzig. Judgment Aggregation in Dynamic Logic of Propositional Assignments. *Journal of Logic and Computation*, 2018.

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* $\boldsymbol{P}$ $\forall$ *alternatives* x, y
($\forall$ *individual* i xP_iy*)*$\to$ $(xw(\boldsymbol{P})y)$

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* $\boldsymbol{P}$ $\forall$ *alternatives* x, y
($\forall$ *individual* i xP_iy*)*$\rightarrow$ $(xw(\boldsymbol{P})y)$

SOLUTION

Introduce a set of situations as "names" for preference profiles:
s is an element of the model domain associated to profile $\boldsymbol{P}^s$.

UN: ($\forall$ *situation* s $\forall$ *alternatives* x, y *(*$\forall$ *individual* i xP_i^sy*)*$\rightarrow$ $(xw(\boldsymbol{P}^s)y)$*)*
It is now almost a first-order sentence.

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

Language

To express statements of this kind we need:

- guards for individuals $I(z)$, alternatives $A(x)$ and situations $S(u)$
- constants a_1, a_2, a_3 for 3 alternatives, plus i_1 and s_1
- a 4-ary relation $p(z, x, y, u)$ to represent the linear order P_z^u of individual z in situation u
- a 3-ary relation $w(x, y, u)$ to represent the social outcome $w(\underline{P}^u)$

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

Language

To express statements of this kind we need:

- guards for individuals $I(z)$, alternatives $A(x)$ and situations $S(u)$
- constants a_1, a_2, a_3 for 3 alternatives, plus i_1 and s_1
- a 4-ary relation $p(z, x, y, u)$ to represent the linear order P_z^u of individual z in situation u
- a 3-ary relation $w(x, y, u)$ to represent the social outcome $w(\underline{P}^u)$

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

Axioms:

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

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

Language

To express statements of this kind we need:

- guards for individuals $I(z)$, alternatives $A(x)$ and situations $S(u)$
- constants a_1, a_2, a_3 for 3 alternatives, plus i_1 and s_1
- a 4-ary relation $p(z, x, y, u)$ to represent the linear order P_z^u of individual z in situation u
- a 3-ary relation $w(x, y, u)$ to represent the social outcome $w(\underline{P}^u)$

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

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 II: Permutation

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

Axioms II: Permutation

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

This property is translated in the next **PERM** axiom:

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

If in situation u the order of individual z is:

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

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 $T_{\textsf{SWF}}$ the axioms presented so far.

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

- **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

$T_{\textsf{ARROW}}$ has no finite models.

First-order: conclusions and other works

Our FO axiomatisation of social welfare functions:

- Other well-known theorem still hold for infinite societies: Sen's impossibility of a Paretian liberal and Kirman-Sondermann theorem translated as "theory T_{AX} has no models"
- (for logic-geeks) the condition of universal domain is not FO-axiomatisable
- We did try using automated provers such as **E** and **Prover 9**: what we could prove SAT could do much faster...

First-order: conclusions and other works

Our FO axiomatisation of social welfare functions:

- Other well-known theorem still hold for infinite societies: Sen's impossibility of a Paretian liberal and Kirman-Sondermann theorem translated as "theory T_{AX} has no models"
- (for logic-geeks) the condition of universal domain is not FO-axiomatisable
- We did try using automated provers such as **E** and **Prover 9**: what we could prove SAT could do much faster...

Researchers from KIT (Germany), Australia, and Denmark used FOL to:

- verify properties of existing voting rules with SMT solvers (example: does STV satisfies proportional representation?)
- compute properties of a specific elections using software-bounded model checking (margin computation: what is the minimal number of misfiled votes to change the result of the election)

Beckert, Bormer, Gore, Kirsten, Schurmann. An introduction to voting rule verification. In *Trends in Computational Social Choice*, U. Endriss (ed), 2017.

Recap and lessons learned for AGAPE

1. The main motivation is different: assist the researcher and eventually discover new theorems. However this kind of formalism could inspire languages for strategic reasoning in auctions.
2. AGAPE seems to be closer to agent-based automated reasoning than theorem proving. However: if an agent faces an auction and wants to know if it's strategy proof, the agent has to prove a theorem!
3. Payoff vs ordinal preferences: the approches described above work very well for preferences as orders or discrete objects. Can we study auctions without money?
4. Universal domain: same as above, to know if something is strategy proof all possible situations must be tested. A general difficulty for this kind of reasoning?

Thank you for your attention!

Other cited references

Ulle Endriss. Logic and Social Choice Theory. In A. Gupta and J. van Benthem, editors, *Logic and Philosophy Today*, volume 2, pages 333-377, College Publications, 2011.

A. Frchette, N. Newman, and K. Leyton-Brown. Solving the station repackingproblem. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2016.

M. B. Caminati, M. Kerber, C. Lange, and C. Rowat. Sound auction specificationand implementation. In *Proceedings of the 16th ACM Conference on Economics and Computation (ACM-EC)*, pages 547564. ACM Press, 2015.