Finding Strategyproof Social Choice Functions via SAT Solving

Christian Geist

COST Meeting, Maastricht, The Netherlands

11 April 2014

Felix Brandt

Christian Geist

DSS
Decision Sciences & Systems

General idea and success cases

- Idea?
  - Reduce to “small” instance (manually using induction)
  - Solve base case on a computer (using universal problem solving approaches such as SAT)
- Successful?
  - Tang/Lin, 2009: Famous impossibilities (Arrow, Sen, Muller-Satterthwaite, etc) for resolute social choice functions
  - G./Endriss, 2011: Automated theorem search (through universal reduction step) for ranking sets of objects
  - (Brandt/G./Seedig, 2014: Finding preference profiles for *k*-majority digraphs)
  - (G., 2014: Finding preference profiles of given Condorcet dimension)
- Today:
  - Method: more evolved technique to also treat strategyproofness for irresolute social choice functions
  - Results: e.g., efficiency and strategyproofness are incompatible for a natural set extension

Results preview and related work

- Two notions of strategyproofness due to Kelly (1977) and Fishburn (1972) (see also Gärdenfors, 1979)
    - Impossibility: Pareto optimality is incompatible with Fishburn-SP
    - Possibility: There is a refinement of BP that is still Kelly-SP
- Closes gaps in the existing (axiomatic) literature on strategyproofness for irresolute social choice functions, e.g.,
    - Kelly (1977)
    - Barberá (1977)
    - Gärdenfors (1979)
    - Ching and Zhou (2002),
    - Brandt (2011)
    - Brandt and Brill (2011)
    - Sanver and Zwicker (2012)

Outline

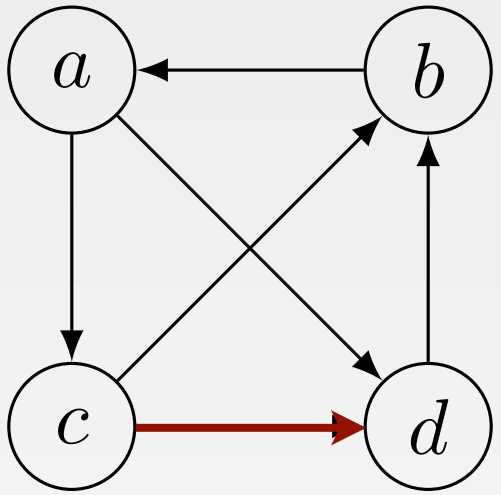
- Preliminaries
  - (Majoritarian) social choice functions
  - Strategyproofness (incl. set extensions of preferences)
- Encoding into SAT
  - Initial encoding
  - Optimizations
- Main results

Preliminaries

- Finite sets of *m* alternatives, *n voters*
  - Voters *i* with complete, anti-symmetric and transitive preference relations R_i over alternatives; strict part P_i (e.g., a P_i b P_i c)
  - Preference profiles R = (R_1, R_2, …, R_n)
- A *social choice function (SCF)* is a function that maps preference profiles to non-empty subsets of alternatives
  - An SCF *f* is *resolute* if $|f(R)|=1$ for all preference profiles *R*
  - An SCF *f* is *neutral* if it treats all alternatives equally
  - An SCF *f* is *majoritarian* if it is neutral and *f*(*R*) only depends on the pairwise majority comparisons of *R* (*majority relation* R_M)
- *Majoritarian SCFs* are also known as *tournament solutions*

Tournament solution examples

- TC(R_M) selects the maximal elements of the transitive closure of R_M
- UC(R_M) consists of all alternatives that are not covered
  - x covers y if $y \, R_M \, v$ implies $x \, R_M \, v$ for all $v \in V$
- BP(R_M) defined based on game theory
  - Alternatives as actions; payoffs based on R_M
  - BP(R_M) consists of all alternatives with positive probability in some Nash equilibrium



TC(R_M)=$\{a,b,c,d\}$

UC(R_M)=$\{a,b,c\}$

BP(R_M)=$\{a,b,c\}$

Allan Gibbard

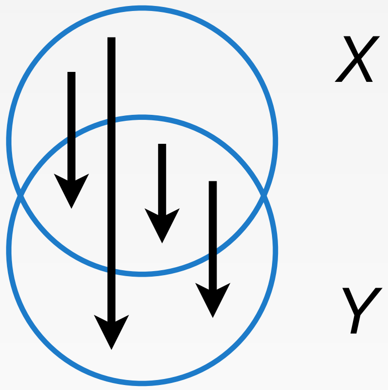
There can be more than one

Mark A. Satterthwaite

- A resolute SCF f is *strategyproof* if there is no R, R', $i \in N$ such that $R_j = R'_j$ for all $j \neq i$ and $f(R')\ P_i\ f(R)$
- Theorem (Gibbard, 1973; Satterthwaite, 1975): Every strategyproof resolute SCF is either imposed or dictatorial.
    - "*[resoluteness] is a rather restrictive and unnatural assumption*" (Gärdenfors; 1976 - a philosopher)
    - "*The Gibbard-Satterthwaite theorem [...] uses an assumption of singlevaluedness which is unreasonable*" (Kelly; 1977 - an economist)
    - "*If there is a weakness to the Gibbard-Satterthwaite theorem, it is the assumption that winners are unique*" (Taylor; 2005 - a mathematician)
- Problem: Resolute SCFs are incompatible with anonymity and neutrality
- Solution: Allow for sets of winners (irresolute SCFs)
    - Natural next question: what preferences do voters have over sets of alternatives

Irresolute SCFs: Kelly's extension

- How to deal with irresoluteness?
  - Assumption: A single alternative is eventually chosen, but the voters do not know *anything* about the tie-breaking mechanism.
  - Under this assumption, the preferences over choice sets are given by Kelly's preference extension $R^K \subseteq A \times A$:
    $X\ R^K\ Y \Leftrightarrow \forall x \in X, y \in Y: (x\ R\ y)$
  - Example
    - $a\ P\ b\ P\ c$ implies that $\{a\}\ P^K\ \{a,b\}\ P^K\ \{b\}\ P^K\ \{b,c\}$
    - $\{a,c\}$ and $\{b\}$ are incomparable
    - $\{a,b\}$ and $\{a,b,c\}$ are incomparable(!)



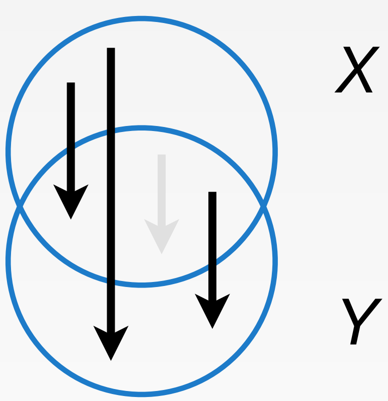
- An SCF f is *P^K-strategyproof* if there is no R, R', $i \in N$ such that $R_j = R'_j$ for all $j \neq i$ and $f(R')\ P_i^K\ f(R)$

What we know about Kelly-strategyproofness

Kelly-strategyproof	*manipulable*
Pareto rule	
Omninomination rule	
Top cycle (TC), 1971	essentially everything else
Uncovered set (UC), 1977	
Minimal covering set (MC), 1988	
Bipartisan set (BP), 1993	

Irresolute SCFs: Fishburn's extension

- How to deal with irresoluteness?
  - Alternative assumption: There is an agent (with preferences *unknown* to the voters) who picks his most preferred alternative from the choice set, e.g., a chairman or one of the voters
  - Under this assumption, the preferences over choice sets are given by Fishburn's preference extension $R^F \subseteq F(U) \times F(U)$:
    $X \, R^F \, Y \Leftrightarrow (\forall x \in X \setminus Y, \, y \in Y: x \, R \, y) \wedge (\forall x \in X, \, y \in Y \setminus X: x \, R \, y)$
  - $X \, R^K \, Y \Rightarrow X \, R^F \, Y$ and hence $R^K \subseteq R^F$
  - Example
    - $a \; P \; b \; P \; c$ implies that $\{a,b\} \; P^F \; \{a,b,c\} \; P^F \; \{b,c\}$
    - $\{a,c\}$ and $\{b\}$ are still incomparable



- An SCF f is *P^F-strategyproof* if there is no R, R', $i \in N$ such that $R_j = R'_j$ for all $j \neq i$ and $f(R') \; P_i^F \; f(R)$

What we know about Fishburn-strategyproofness

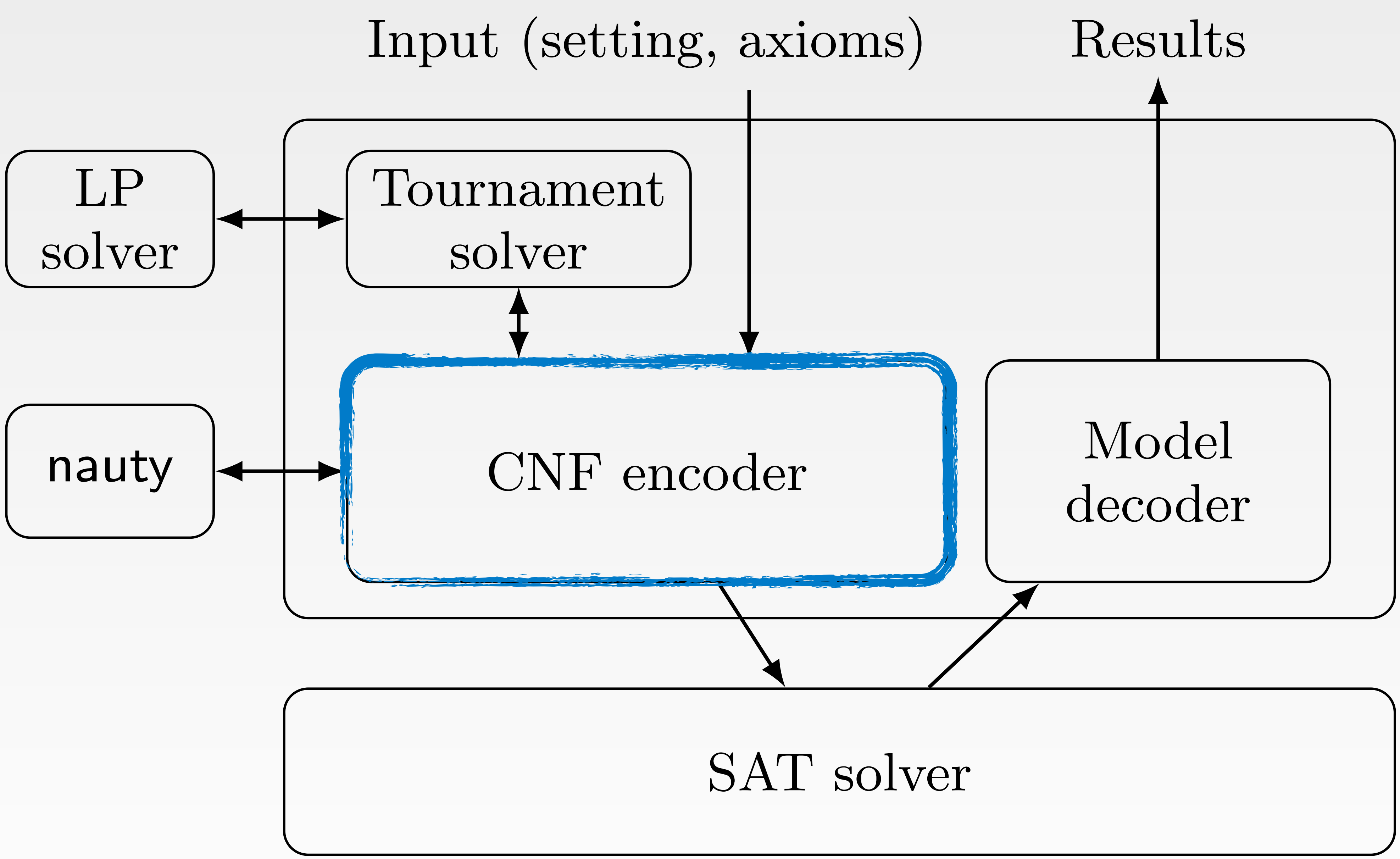
Fishburn-strategyproof	*manipulable*
Pareto rule	essentially everything else
Omninomination rule	
Top cycle (TC), 1971	

Logically equivalent but simpler: tournament-strategyproofness

- A majoritarian SCF f is said to be P^E-tournament-strategyproof if there are no T, T' and $P_\mu \supseteq (T - T')$ such that $f(T')\ P_\mu^E\ f(T)$
- Lemma. A majoritarian SCF is P^E-strategyproof iff it is P^E-tournament-strategyproof
- Enables more efficient check on a computer, but still large

Alternatives	**4**	**5**	**6**	**7**
Choice sets	15	31	63	127
Tournaments	64	1,024	32,768	~ 2
Unlabeled tourn.	4	12	56	456
Majoritarian SCFs	**50,625**	**~ 10**	**~ 10**	**~ 10**

High-level system architecture



Basic encoding: goal and variables

- Goal: Encode full problem (of fixed size) into SAT
  - Find propositional formula that is satisfiable iff base case is true
- Variable symbols $c_{T,X}$ to represent $f(T) = X$
- Explicit axioms
  - (Tournament-)strategyproofness
  - Pareto optimality
  - …
- Context axioms
  - Functionality (of the choice function)
  - Neutrality

Basic encoding: example axiom

- Apart from explicit axioms, 2 main contextual axioms
  - Functionality (of the choice function)
  - Neutrality
- Example: Functionality

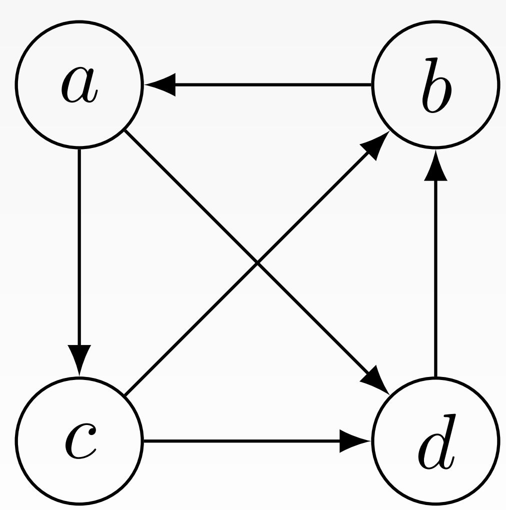
$$(\forall T)\ ((\exists X)\ c_{T,X} \wedge$$
$$(\forall Y, Z)\ Y \neq Z \to \neg(c_{T,Y} \wedge c_{T,Z}))$$

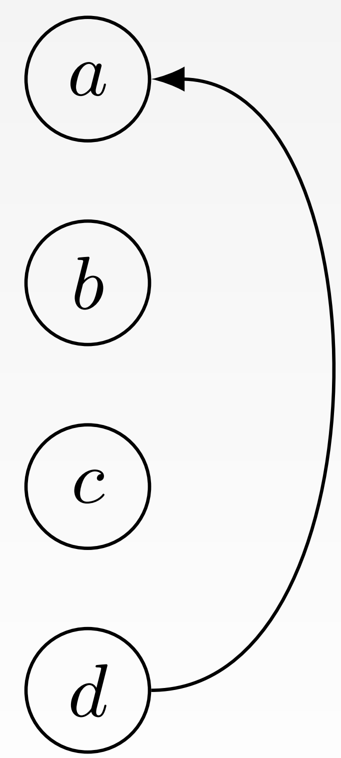
$$\equiv \bigwedge_{T} \left(\left(\bigvee_{X} c_{T,X} \right) \wedge \bigwedge_{Y \neq Z} (\neg c_{T,Y} \vee \neg c_{T,Z}) \right)$$

foreach *Tournament T* **do**
  foreach *Set X* **do**
    variable(c(T, X));
  end
  newClause;
  foreach *Set Y* **do**
    foreach *Set* $Z \neq Y$ **do**
      variable_not(c(T, Y));
      variable_not(c(T, Z));
      newClause;
    end
  end
end

Neutrality is not as innocent as it seems

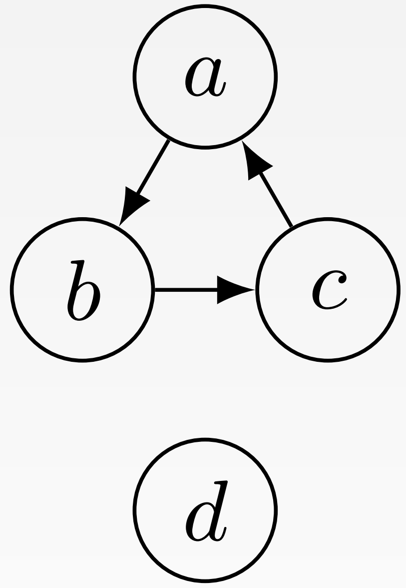
- Formally:
  $\pi(f(T)) = f(\pi(T))$ for all tournaments T and permutations π
- Has implications across tournaments and even on single tournaments
  - Across: Isomorphic tournaments
    - *canonical tournaments T_C*





Neutrality is not as innocent as it seems

- Formally:
  $\pi(f(T)) = f(\pi(T))$ for all tournaments T and permutations π
- Has implications across tournaments and even on single tournaments
  - Across: Isomorphic tournaments
    - *canonical tournaments T_C*
  - Within: Orbits



Neutrality is not as innocent as it seems

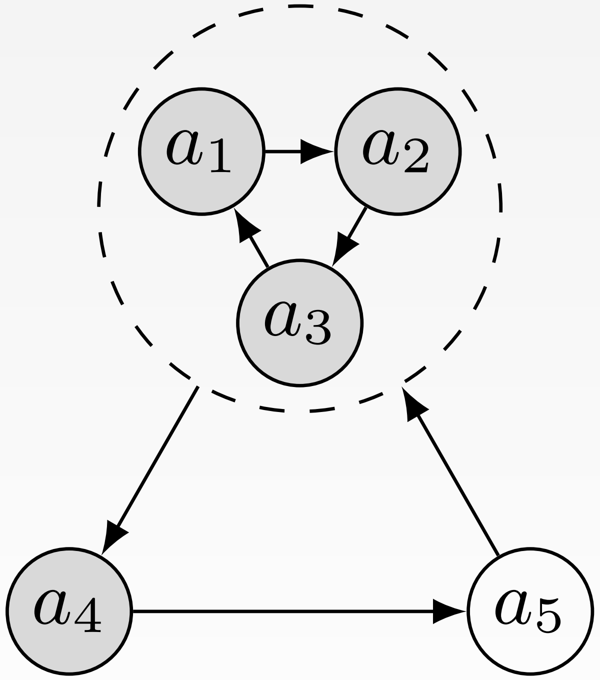
- Formally:
  $\pi(f(T)) = f(\pi(T))$ for all tournaments T and permutations π
- Has implications across tournaments and even on single tournaments
  - Across: Isomorphic tournaments
    - *canonical tournaments T_C*
  - Within: Orbits
- Lemma. *Neutrality* is equivalent to the conjunction of
  - Canonical isomorphism equality: $f(T) = \pi_T(f(T_C))$
  - Orbit condition: $O \subseteq f(T_C)$ or $O \cap f(T_C) = \emptyset$
- Further optimizations of the encoding are possible

Main result: Pareto-optimility and Fishburn-strategyproofness are incompatible

- A SCF is Pareto-optimal if its choice sets never contain a Pareto dominated alternative
- Theorem. For $m \geq 5$ there is no majoritarian SCF f that satisfies Fishburn-strategyproofness and Pareto-optimality
  - Lemma. *Pareto-optimality* ⇔ *refinement of UC*
  - Lemma. Base case $m = 5$: automatic verification
    - Fishburn-strategyproofness
    - Refinement of UC
  - Lemma. ∃ strategyproof maj. SCF $f \subseteq UC$ for m+1 alternatives ⇒ ∃ strategyproof maj. SCF $f' \subseteq UC$ for m alternatives

Positive result: Kelly-SP

- Theorem. There exists a refinement of BP which is still Kelly-strategyproof
  - BP is not the smallest majoritarian SCF satisfying Kelly-strategyproofness
  - The only strategyproof refinement on 5 alternatives
  - Not all desirable properties of BP carry over
- Defined like BP with the exception of:



Proof extraction is possible

Proof trace

```
p cnf 372 17
c Orbit
248 0
c C5
280 293 294 295 296 309 310 0
c 3-cycle-in-3-cycle
c Refinement of UC
1 0
94 95 96 101 102 103 104 0
c Strategyproofness (Fishburn)
-94 -94 0
-95 -95 0
-101 -101 0
-102 -102 0
-96 -1 0
-103 -1 0
-248 -295 0
-248 -309 0
-293 -1 0
-294 -1 0
-280 -1 0
-296 -1 0
-310 -104 0
0
```

Minimal
UNSAT core

Proof. Let f
and assume
strategyproo
ers $C \subseteq N$
$R_j = R'_j$ fo

Human-readable
proof

We successfully transferred SAT-based theorem proving to irresolute majoritarian social choice functions

- (Brief) introduction to
  - Irresolute SCFs
  - Majoritarian SCF (tournaments rather than preference profiles)
  - Kelly-/Fishburn-strategyproofness
- Encoding
  - Contextual and explicit axioms ($c(T,X)$)
  - Optimization techniques for improved performance
- Initial new results
  - Incompatibility of Pareto-optimality and Fishburn-strategyproofness
  - Kelly-strategyproof refinement of BP
- (Semi-automatic) proof extraction
- Universality and ease of adaptation most likely to enable further results