An economist's guide to mechanized reasoning or My computer just proved 84 impossibility theorems

Kerber, Manfred; Lange-Bever, Christoph; Rowat, Colin

Document Version
Peer reviewed version

Citation for published version (Harvard):
Kerber, M, Lange, C & Rowat, C 2012, 'An economist's guide to mechanized reasoning or My computer just proved 84 impossibility theorems' Initiative for Computational Economics summer school, Chicago, United States, 25/07/12, .

An economist's guide to mechanized reasoning *or* My computer just proved 84 impossibility theorems

Manfred Kerber[1] Christoph Lange[1] Colin Rowat[2]

[1]Computer Science, University of Birmingham
[2]Economics, University of Birmingham

July 2012

Initiative for Computational Economics, University of Chicago/ANL

EPSRC grant EP/J007498/1

Overview

1. What has mechanized reasoning achieved?
2. Uses in economics
   - Arrow's impossibility theorem
   - Other economic applications
3. Worked example: pillage games
4. Possible next steps in economics
   - Promising problem domains
   - A source of metrics?
5. Resources for economists
6. Conclusions

The four color map problem

- 1852: Francis Guthrie stumps de Morgan, who adopts the problem
- 1879: Alfred Kempe proposes proof; the Royal Society only discovers its errors a decade later
- 1976: assembly language code written to 'prove' result [AH77; AHK77]
  - graph theory identifies
    1. reducible configurations
    2. a set of $< 2{,}000$ minimal possible counter-examples
  - computer searches over these minimal examples
  - impossible to hand-check the whole proof (over 400 pages), minor errors surfaced [AH89, p.23], doubts remained
- [Gon08]: formalized whole proof as a program for evaluation in Coq proof system

Robbins problem: bases for Boolean algebras

- Robbins (1930s): for any Boolean algebra, are the following equivalent:

  (HUN) $\overline{\overline{X} \vee Y} \vee \overline{\overline{X} \vee \overline{Y}} = X$

  (ROB) $\overline{\overline{X \vee Y} \vee \overline{X \vee \overline{Y}}} = X$

  ▸ trivial with single atom, $\mathcal{E} = \{a\}$, so that $X, Y \in \{0, 1\}$, but . . .

- question came at a period of intense interest in the axiomatic foundations of logic
- beguilingly simple, but open question for 60 years, and favorite of Tarski [HMT71, p.245]
  - little intuition: only example of Robbins algebra was also Boolean
- [McC97] at Argonne exploits Winker's sufficient conditions
  - automated first order logic solver, EQP, generates proof in 8 days, using 30MB memory
  - 17 step proof, after trying 17,666 (complex) steps; humans can check it
  - fine-tuning produces an 8 step proof in 5 days

Stacking cannon balls: the Kepler conjecture

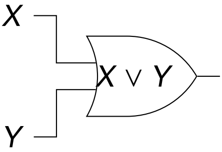
- 1611: Kepler conjectures that face-centred cubic packing of spheres achieves maximum density
- 1900: Hilbert includes it in problem 18 of his 23 unsolved problems
- 1953: Tóth proves that finitely many calculations could check all cases

(http://tinyurl.com/3bxx2t)

- Hales implements Tóth, minimizing an 150 variable function for 5,000 cases
  - solved 100,000 linear programming problems
  - [Hal05]: submitted in 1998; by 2003, 12 referees were "99% certain" was correct, but "will not be able to certify it . . . because they have run out of energy to devote to the problem"
- 2003: Hales launches *Project FlysPecK*, using prover HOL Light, to formally prove; expected to take 20 years

Hardware, software verification: $\frac{4{,}195{,}835}{3{,}145{,}727} \approx 1.33374$?

- Pentium floating point division bug (1994): worst known relative error 0.006%; few affected, but costs Intel $475mn
- destruction of Ariane 5 Flight 501 (1996): 64-bit floating point value converted to 16-bit signed integer value



1. model hardware, software systems as logical
2. prove theorem for each IEEE property to be implemented
   - e.g. sufficient condition for perfect square root rounding is

$$\left|\sqrt{a} - s^*\right| < \left|\sqrt{a} - m\right| \forall a \in \mathbb{R}$$

   where algorithm returns s^*, m is the midpoint between the bounding floats [Har06]
3. model checking at ATP end of spectrum, more common; theorem proving at ITP end, less common [Woo+09]

Eureqa: deducing Newton's laws [SL09]?

Distilling Free-Form Natural Laws from Experimental Data (YouTube video)

> *We're going to see scientific results that are correct, that are predictive, but are without explanation. We may be able to do science without insight, and we may have to learn to live without it. Science will still progress, but computers will tell us things that are true, and we won't understand them. (Steven Strogatz, 2010, NYT)*

Watson beats the humans on Jeopardy

IBM's Watson supercomputer destroys all humans in Jeopardy (YouTube video)

- probabilistic expert system capable of
  - natural language reasoning
  - case-based reasoning (q.v. [GS01])
- v. Deep Blue: broad knowledge rather than narrowly specialized
- now signed up with Citigroup

How can mechanized reasoning help economics?

When possible, shall illustrate with Arrow's impossibility theorem:

- F formal representation and retrieval
  - searching for $a^2 + b^2 = c^2$ finds $x^2 = y^2 + z^2$ [KMP12]
- H makes hidden assumptions explicit
  - [Gea01; Gea05; Nip09]
- ∃ confirms existing results
- C cleans up proofs
- S suggests new proof strategies
- N helps find new results (inc. new types of results)
  - [TL09; GE11]
- R helps review work
  - [KRW11]

A checklist

1. a tractable problem
   - are there a finite number of finite cases to consider (maybe with an induction step)?
   - n.b. [CHH02]: Deep Blue v. Kasparov in 1997 usually searched 6 – 16 ply deep, with max 40 ply
2. an appropriate logic (and calculus) for handling your conjectures
   1. provers don't compromise on soundness
      - if it is deduced, it is a property: $(\Gamma \vdash \varphi)$ then $(\Gamma \models \varphi)$
      - trivial soundness: "on the advice of counsel, I respectfully assert . . . "
   2. expressiveness: must be able to formulate all relevant properties
   3. completeness: any question asked can, in principle, be answered by skillful use of the logic's calculus
      - if it is a property, it can be deduced: $(\Gamma \models \varphi)$ then $(\Gamma \vdash \varphi)$
   4. decidable: if an answer exists, there is an algorithm for deriving it

   - art: trading off expressiveness, completeness and decidability
3. a solver that efficiently implements the calculus

A brief word on classical logics

expressiveness ↓

propositional: concrete, finite statements
"Ken is a dictator over pair {SITE, ICE}"
sound, complete, decidable
not expressive
Chaff; [TL09]

first order: propositional + quantification ($\forall, \exists$) over objects
"there exists a dictator, n, over any pair $\{a, b\}$"
sound, complete, more expressive (Gödel completeness)
not decidable
Prover9, Vampire, Prolog; [GE09]

higher order: FOL + quantification over functions, predicates
"if n is an X over $\{a, b\}$ then n is an X over all pairs"
sound, very expressive
not complete (Gödel incompleteness) or decidable
HOL Light, Isabelle; [Har06]
(n.b. FOL + set theory replicates HOL
uses sets to define functions, predicates; e.g. Mizar; [Wie09])

Caveat

the expectation was that these advances [in automated reasoning] would also have significant impact on the practice of doing mathematics. However, so far, this impact is small. We think that the reason for this is the fact that automated reasoning so far concentrated on the automated proof of individual theorems whereas, *in the practice of mathematics, one proceeds by building up entire theories in a step-by-step process. This process of exploring mathematical theories consists of the invention of notions, the invention and proof of propositions (lemmas, theorems), the invention of problems, and the invention and verification of methods (algorithms) that solve problems. [Buc06]*

Arrow's impossibility theorem

A constitution respects `UN` *if society puts alternative a strictly above b whenever every individual puts a strictly above b. The constitution respects* `IIA` *if the social relative ranking (higher, lower, or indifferent) of two alternatives a and b depends only on their relative ranking by every individual. The constitution is a* `D` *by individual n if for every pair a and b, society strictly prefers a to b whenever n strictly prefers a to b. [Gea05]*

Theorem (Arrow [Gea05])

(For two or more agents, and three or more alternatives,) any constitution that respects transitivity, `IIA`*, and* `UN` *is a* `D`*.*

Social choice theory turns out to be perfectly suitable for mechanical theorem proving. . . . However, it is unclear if this will lead to new insights into either social choice theory or theorem proving. [Nip09]

we form an interesting conjecture and then prove it using the same [mechanized] techniques as in the previous proofs. . . . the newly proved theorem . . . subsumes both Arrow's and Wilson's theorems. [TL09]

When applied to a space of 20 principles for preference extension familiar from the literature, this method yields a total of 84 impossibility theorems, including both known and nontrivial new results. [GE11]

Geanakoplos' three brief proofs

	1st	3rd
hand	1 page	1 page
Isabelle [Nip09]	350 lines (6 pages)	300 lines
Mizar [Wie07; Wie09]	1100 lines	

Geanakoplos' proofs [Nip09]

1st proof [Gea01; Nip09]

- statement in extremal lemma required 20 line auxiliary proof
- equation of pivotal and dictator only hinted at originally

3rd proof [Gea01; Nip09]

- pairwise neutrality lemma mentions two profiles not explicitly considered
- minor missing case in pairwise neutrality lemma
- therefore, could not formalize proof

Nipkow e-mails Geanakoplos in 2002; both proofs revised in [Gea05]

Restating Arrow's theorem [GE09]

- surprised that could mostly formalize Arrow in FOL, as T_{ARROW}
  - quantification over preference profiles feels second order

Theorem (Arrow à la [GE09])

T_{ARROW} has no finite models.

∴ no counterexamples to T_{ARROW} for finite # of agents, alternatives

> *We designed a step-by-step proof . . . for 2 individuals and 3 alternatives . . . At each step we received a negative response, with the prover exceeding the search space limits or not providing an answer in a reasonable amount of time.*

- prover seemed unable to apply "permutation, guessing the correct sequence of swaps to get from a profile to another"
- an automated proof exists as FOL is complete, but "for every finite number of individuals there is a (possibly different) first-order proof"

Inducing Arrow from 2 agents, 3 alternatives [TL09]

- manually: prove induction from 2 agent, 3 alternative base case
  - [Suz00]: induction proof from 2 agent, *n* alternative base case

1. constraint satisfaction problem
   - not feasible to generate all $6^{36} \approx 10^{28}$ SWF on the base case
   - find all SWF satisfying `U` and `IIA`, and verify are `D`
   - CSP: $\langle X, D, C \rangle$, where
     - X is set of variables ($36 = 6 \times 6$ preference profiles)
     - D is their domain (6 linear orderings for each profile)
     - C is the constraint set (`U` and `IIA`)
   - Prolog code: finds 2 dictatorial SWF in < 1 second on desktop
2. Boolean satisfiability problem (SAT) $\subset$ CSP
   - express clauses of Boolean variables in conjunctive normal form

$$(x \vee y \vee z) \wedge (\neg x \vee \neg y \vee z) \wedge (y \vee \neg y) \wedge \cdots$$

   - encoded base case in 35,973 variables in 106,354 clauses
   - situation calculus: `swap` action augments propositional logic to permutate profiles ▸ encoding axioms
   - Chaff2 SAT solver shows inconsistency, < 1 second on desktop
     - SAT solver: theorem prover in propositional logic

New theorem generalizes Arrow's, Wilson's

- only 94 of 6^{36} base case SWFs satisfy `IIA`
- their inspection establishes the base case for a new theorem

Theorem ([TL09])

If a social welfare function W on (N, A) satisfies `IIA`, then for every subset Y of A such that $\|Y\| = 3$,

1. *W_Y is dictatorial, or*
2. *W_Y is inversely dictatorial, or*
3. *The range of W_Y has at most 2 elements, whose [Kendall tau] distance is at most 1.*

- a new induction lemma then establishes the theorem

Ranking sets of objects [KP84; BBP04]

- [GE11]: instead of an induction lemma per base case, a broadly applicable induction theorem for ranking sets of objects
- many-sorted logic for set preferences (MSLSP), a first order logic, allows separate quantification over elements, sets
- using 20 known axioms [BBP04], use SAT solver to generate 84 impossibility theorems from c.1mn combinations
  1. yields new theorems
  2. strengthens existing theorems
  3. aids understanding of axioms' role
     - first [?] impossibility result without either `GF` or `SDom`
     - `LIN` appears in all theorems; `evenExt`, `REFL` occur in none; `intIND` occur in all for 7 or 8 choice items, and never for fewer than 5
  4. establishes suspected results
     - [BPX00] `SDom`, `IND`, `SUAv` and `STopMon` characterize the min-max ordering
     - [Arl03] $n = 5$ counter-example: min-max ordering violates `IND`
     - [GE11] impossibility with ≥ 4 choice items, including manual proof

Open questions

1. extend [GE11] beyond $n = 8$?
2. ★ apply [GE11] to other axiomatic social choice/utility theory?
   - e.g. judgment aggregation [LP10]
3. [GE11] can't express `NEU` neutrality axiom in MSLSP. Is there another (tractable) logic that can?
   - see [ÅHW09; ÅHW11] on developing logics

Unique PNE payoffs in 2 agent games [TL11a]

1. express properties, φ, of 2 agent games, $\Gamma_{2\times2}$, in first order logic e.g. strictly competitive games' weakly opposed preferences
   - for all $\boldsymbol{a}, \boldsymbol{a}' \in A_1 \times A_2$: $\boldsymbol{a} \succsim_1 \boldsymbol{a}' \equiv \boldsymbol{a}' \succsim_2 \boldsymbol{a}$, an example of

$$(l_1 \vee l_2) \wedge (l_3 \vee l_4) \qquad (1)$$

     where each l_j is either $\boldsymbol{a} \succsim_i \boldsymbol{a}'$ or its negation, $\neg(\boldsymbol{a} \succsim_i \boldsymbol{a}')$
   - known to have unique PNE payoffs

$$NE(\boldsymbol{a}) \wedge NE(\boldsymbol{a}') \supset (\boldsymbol{a} \sim_1 \boldsymbol{a}') \wedge (\boldsymbol{a} \sim_2 \boldsymbol{a}') \qquad (2)$$

2. manual proof: iff a counterexample to (2) exists for a class of games defined by a sentence like (1), it exists for a 2×2 game
3. generate all $2^4 \times 15^4 = 810{,}000$ properties, φ, of form (1) and all $75^2 = 5{,}625$ 2×2 games, $\Gamma_{2\times2}$
4. for each φ, test whether $(\Gamma_{2\times2} \vDash \varphi \supset (2))$ on all $\Gamma_{2\times2}$
5. for all that do, prune to collect the weakest
   - complication: logical entailment generally not decidable in FOL

Unique PNE payoffs in 2 agent games: results

- find three types of (weakest) uniqueness conditions with form 1:
  1. weakly unilaterally competitive (WUC) [KT92] $\supset$ strictly competitive: proves weakest uniqueness condition of form 1
  2. 1's self-interest helps 2, while 2's hurts 1: $\frac{\partial u_2}{\partial u_1} > 0$, $\frac{\partial u_1}{\partial u_2} < 0$
  3. both agents can simultaneously achieve their maximal payoffs
- in strict games, $\Gamma_s \subset \Gamma$, in which each profile has a distinct payoff,

$$(\boldsymbol{a} \gtrsim_i \boldsymbol{a}' \gtrsim_i \boldsymbol{a}) \Rightarrow (\boldsymbol{a} = \boldsymbol{a}')\,,$$

  so that unique PNE payoffs imply unique PNE:
  1. new: weakly unilaterally competitive for player i (if WUC for both, then WUC)
  2. games with dominant strategies
  3. a condition that cannot be satisfied in games larger than 3×3
- [TL11b] manually proves characterization results for two classes of games suggested by mechanized work

Open questions

1. ★ standard uniqueness conditions for PNE include
   - dominant diagonal in supermodular games [MR90]: complete lattices, upper semi-continuity, second partial derivatives
   - unique (possibly mixed) Nash equilibrium iff both players have same number of strategies in support of their BR functions [Kre74]

   1. can the above uniqueness conditions be formalized as conjunctions of the two [TL11a] binary clauses? If not, what is their simplest formalization?
   2. consequences of relaxing requirement that $\geq_1, \geq_2$ alternate in two binary clauses?
   3. are conjunctions of three binary clauses tractable?
   4. how easily derive [TL11a] results beyond $n = 2$?

Reasoning about coalitional games [ÅHW09]

- present logics for handling NTU cooperative games
  1. coalitional game logic: expressive, but only for finite games
     - 21 line proof that the core is a subset of any stable set
  2. modal coalitional game logic: less expressive for finite games, but can handle infinite games
     - 9 line proof that the core is a subset of any stable set
- analyze soundness, completeness, model checking, satisfiability

Pillage games [Jor06]

- richer than characteristic, partition function forms
- n agents split a unit pie, $\sum_{i=1}^{n} x_i = 1, \boldsymbol{x} \in \boldsymbol{X}$
- dominance relation, $\succ$, represented by power function, π, increasing in coalitional membership, resources

  (WC) $C \subset C' \subseteq I \Rightarrow \pi(C', \boldsymbol{x}) \geq \pi(C, \boldsymbol{x}) \, \forall \boldsymbol{x} \in X$

  (WR) $y_i \geq x_i \forall i \in C \subseteq I \Rightarrow \pi(C, \boldsymbol{y}) \geq \pi(C, \boldsymbol{x})$

  (SR) $C \neq \emptyset \subseteq I$ and $y_i > x_i \forall i \in C \Rightarrow \pi(C, \boldsymbol{y}) > \pi(C, \boldsymbol{x})$
- often analytically convenient if π also satisfies

  (AN) anonymity: if $\sigma : I \rightarrow I$ is a 1:1 onto function permuting the agent set, $i \in C \Leftrightarrow \sigma(i) \in C'$, and $x_i = x'_{\sigma(i)}$ then $\pi(C, \boldsymbol{x}) = \pi(C', \boldsymbol{x}')$
- $\boldsymbol{x}$ dominates $\boldsymbol{y}$ (written $\boldsymbol{x} \succ \boldsymbol{y}$) iff $\pi(W, \boldsymbol{y}) > \pi(L, \boldsymbol{y})$, where

$$W \equiv \{i | x_i > y_i\} \text{ and } L \equiv \{i | y_i > x_i\}$$

Lemmas 1 and 2 [KR09]

Lemma

Any power function, $\pi(C, \boldsymbol{x})$, can be represented by another, $\pi'(C, \{x_i\}_{i \in C})$, which depends only on the resource holdings of its coalition members.

Proof.

Consider arbitrary $\boldsymbol{x}, \boldsymbol{y}$ such that $x_i = y_i \forall i \in C \subseteq I$. Then $y_i \geq x_i$ and $x_i \geq y_i$ so that axiom WR requires $\pi(C, \boldsymbol{y}) \geq \pi(C, \boldsymbol{x}) \geq \pi(C, \boldsymbol{y})$. For this to hold, $\pi(C, \boldsymbol{x})$ cannot depend on x_j for any $j \notin C$. □

Lemma

Let $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$ such that $W = \{i | y_i > x_i\} = \{1\}$ and $L = \{i | y_i < x_i\} = \{2\}$. Then, for any power function satisfying axiom AN, $\boldsymbol{y} \succ \boldsymbol{x} \Leftrightarrow x_1 > x_2$.

Encoding the lemmas in *Theorema* [KRW11]

a reasonable rule of thumb when formalizing is that it takes about one week of full time work to formalize a textbook page. [Wie09]

- typed v. untyped?
- procedural v. declarative?
- compute v. prove [Gon08, p.1385]?

1. Lemma 1
   - predicate logic ($\approx$ FOL) prover generates proof automatically
   - even ATP required good knowledge of the proof
   - 10 page human-readable proof for full search
   - 5 page proof for search settings used in [KRW11]
   - 3 page proof when automatically tidied to leave only steps in final argument (de Bruijn factor $\approx$ 25?)
2. Lemma 2
   - set theory ($\approx$ FOL+SET) prover invoked
   - needed guidance (ITP), partly as permutation is hard [GE09]: we assert auxiliary lemmas

Open questions

1. we would like to extend pillage games results in [KR09] by dropping the anonymity axiom (would ease empirical tests)
2. ★ establish minimal counter-examples to the existence of stable sets? See [Luc68b; Luc68a; LR82] for the counter-examples of record.

> *Finding stable sets involves a new tour de force of mathematical reasoning for each game or class of games that is considered. Other than a small number of very elementary truisms . . . there is no theory, no tools, certainly no algorithm . . . you just have to slug it out anew every time. And because stable sets do not always exist, you cannot even be sure that you are looking for something that is there. [Aum85]*

3. extend [ÅHW09] to TU games?

Mechanism design and auction theory

- social choice, mechanism design, cooperative game theory all structurally similar, rely on axiomatic methods [Suz02]
  - social choice: given agent types, which SCF satisfy axiom set?
  - mechanism design: given SCF, can designer recover types from messages, and implement SCF via a transfer function?
    - Gibbard-Satterthwaite: [Gib73] proof uses Arrow; [Sat75] is direct
  - cooperative games
    - binary relation is $\multimapinv$ rather than $\succsim_i$
    - SAT solving (e.g. for model checking) often uses BDD algorithm on acyclic digraph
- auction theory as a subset
  - model checking important given sums involved (similarly with matching problems)
  - in combinatorial auctions, revenue-maximizing design is *NP*-complete even with one bidder [CS04]
  - sophisticated auctions often run ‘in the wild’ with few formal results Klemperer [Kle10]
    - how analyze bidders seeking to borrow at up to 5% against £80 mn strong collateral, at up to 7% against £100 mn weak collateral, and is willing to pay anything to borrow £40 mn?

Econometrics software

it is not safe to assume that econometric software is accurate [McC09]

package	μ	α_0	α_1	β_1
X1	-0.00540	0.0096	0.142	0.821
X2	-0.00608	0.0098	0.144	0.818
X3	-0.00624	0.0108	0.153	0.806
X4	-0.00619	0.0108	0.152	0.806
X5	-0.00613	0.0107	0.153	0.806
X6	-0.00919	0.0098	0.144	0.818
X7	-0.00619	0.0108	0.153	0.806

GARCH estimates pre-FCP benchmark [McC09]

- user errors: [LL82] corrects [Fel74]; [FG05] corrects [DL01]
- best practice seems to be
  1. identify stable algorithms, including by functional testing on certified value (à la NIST datasets)
  2. formally prove that they are correctly encoded (à la Formal Linear Algebra Methods Environment [Gun+01; Bie+05])
- [McC10] advocates using open-source R rather than a black-box

Finance and risk management

There is no faster way for a trading firm to destroy itself than to deploy a piece of trading software that makes a bad decision over and over in a tight loop. [Min11]

1. finance second largest use domain, after transport [Woo+09]
   - largely model checking of transactions processing software in distributed domains
   - e.g. φ include: “no value is created”, “all value is accounted for”
2. functional programming increasingly used in finance
   - harder to write, but ‘more correct’ once written
   - as avoids ‘side effects’ (only returns result; doesn’t alter global variables, read or write data, . . .) only need to verify routines once
   - develops from 1930s’ formal system, Church’s λ-calculus
   - theorem provers like Coq, HOL Light written in OCaml
   - R, Mathematica support functional programming
   - Credit Suisse “develops and maintains mathematical models to manage derivatives trading and analyze investment portfolios” in F#

Is Basel's market risk management meaningful?

[JP Morgan] adopted a new VAR model . . . only to switch back . . . after losses spiralled – the old model showed the unit's $129 million average [Q1] VAR . . . was almost twice as high as the $67 million the bank had publicly reported. . . .

But a former senior regulator at the OCC says . . . "The OCC . . . validates the framework by which the institutions construct and validate their own models . . . it is not possible to dig deeply into each model." [Car12b]

"The cynical view is that the [traders] figure out the weaknesses in the new VAR model, and put on positions that do not result in increased modelled risks," says Christopher Finger . . . at MSCI . . .

"The new VAR model was data-mined to produce the desired number of breaches in the past, and . . . halved the VAR relative to the old model . . . " says the chief risk officer at one hedge fund. . . .

A senior risk manager at a US bank puts it more bluntly: "This isn't because of a modelling problem . . . This was a fundamental failure of high-level risk management" [Car12a]

Open questions

1. which logics are capable of encoding product-mix auction bids [Kle10]?
2. encode coherent [Art+99] and spectral [Ace02] risk axioms, results?
3. connect finance literature on axiomatic risk management [Art+99; Ace02] to utility theory's axiomatisations of ambiguity [ES10]
4. how to check large internal VaR models?
   - hedge funds' operational risk measures currently seem related to disclosure statements [Bro+08; Bro+09]
   - ditto algo trading models: flash crashes, floating point representation [Har06]

A metric for bounded rationality

- have considered MR as a tool for solving problems
- but may also provide a variety of metrics for bounded rationality
- thus, alternative to existing approaches

1. finite automata
   - can't encode strategies for which might need to count infinitely high
   - e.g. "punish n^{th} deviation for n periods" [Rub98]
2. level-k reasoning [CCGIrt]
   - L_0 as naïve play, L_1 as BR to $L_0, \ldots, L_{n+1}$ as BR to $L_n \ldots$
   - best response mapping may be arbitrarily complex, but reasoner doesn't know induction
   - thus, Bertrand duopoly, traveler's dilemma: level-k converges very slowly

- by contrast, higher order logic allows induction over the natural numbers, allowing modeling of important computational processes

ForMaRE: Formal Mathematical Reasoning in Economics

A hub for MR/ATP within economics, including:

1. a wiki containing
   1. project pages for all known applications of MR/ATP to economics, including links to their code
   2. a list of 100 theorems in economics, containing (so far) 50 theorems, of which 7 have been formalized
2. a discussion list

Other general MR/ATP resources include:

1. Sutcliffe and Suttner's TPTP Problem Library for Automated Theorem Proving
   - CADE ATP System Competition (CASC)
2. Verified Software Repository: presently not well developed
3. Wiedijk
   - $\sqrt{2}$ is irrational in 17 different provers
   - reciprocal of power series challenge at ICMS 2006, inc. demo videos for Isabelle, HOL Light, Mizar, ProofPower, Coq; re-worked by Felix Breuer

1. MR has solved open problems in specific areas of mathematics
2. exploiting these powerful techniques will require new skills
   - most powerful mechanized/formal results may require manually establishing new lemmas, theorems
3. if inappropriately applied, cumbersome and almost useless
4. MR broader than 'just' theorem proving: F, H, ∃, C, S, N, R
5. ITP more successful, more broadly applicable than ATP
   - ATP like a driverless car; ITP helps the driver
6. Moore's law may only add a ply or two to search depth in the near future: thus, not explosive progress in theorem proving since 1997
7. recent surge of serious effort within computer science to apply formal methods to economics, largely outside our awareness

References I

[Ace02] Carlo Acerbi. "Spectral measures of risk: a coherent representation of subjective risk aversion". In: *Journal of Banking and Finance* 26.7 (2002), pp. 1505–1518.

[AH77] Kenneth Appel and Wolfgang Haken. "Every Planar Map is Four Colorable Part I: Discharging". In: *Illinois Journal of Mathematics* 21.3 (1977), pp. 429–490.

[AH89] Kenneth Appel and Wolfgang Haken. *Every planar map is four colorable*. American Mathematical Society, 1989.

[AHK77] Kenneth Appel, Wolfgang Haken, and John Koch. "Every Planar Map is Four Colorable Part II: Reducibility". In: *Illinois Journal of Mathematics* 21.3 (1977), pp. 491–567.

[Arl03] Ritxar Arlegi. "A note on Bossert, Pattanaik and Xu's "Choice under complete uncertainty: axiomatic characterization of some decision rules"". In: *Economic Theory* 22.1 (2003), pp. 219–225.

[Art+99] Philippe Artzner et al. "Coherent measures of risk". In: *Mathematical Finance* 9.3 (1999), pp. 203–228.

References II

[Aum85] Robert J. Aumann. "What is game theory trying to accomplish?" In: *Frontiers of Economics*. Ed. by Kenneth Arrow and Seppo Honkapohja. Basil Blackwell, 1985.

[BBP04] Salvador Barberà, Walter Bossert, and Prasanta K. Pattanaik. "Ranking sets of objects". In: *Handbook of Utility Theory*. Ed. by Salvador Barberà, Peter J. Hammond, and C. Seidl. Vol. II. Dordrecht: Kluwer Academic Publishers, 2004, pp. 893–977.

[Bie+05] Paolo Bientinesi et al. "The science of deriving dense linear algebra algorithms". In: *ACM Transactions on Mathematical Software* 31.1 (2005), pp. 1–26.

[BPX00] Walter Bossert, Prasanta Pattanaik, and Yongsheng Xu. "Choice under complete uncertainty: axiomatic characterizations of some decision rules". In: *Economic Theory* 16.2 (2000), pp. 295–312.

[Bro+08] Stephen Brown et al. "Mandatory Disclosure and Operational Risk: Evidence from Hedge Fund Registration". In: *Journal of Finance* 63.6 (2008), pp. 2785–2815.

References III

[Bro+09] Stephen Brown et al. “Estimating Operational Risk for Hedge Funds: The Omega-Score”. In: *Financial Analysts' Journal* 65.1 (2009), pp. 43–53.

[Buc06] Bruno Buchberger. “Mathematical Theory Exploration”. In: *IJCAR*. 2006, pp. 1–2.

[Car12a] Laurie Carver. “JP Morgan's ‘London whale’ losses spark VAR debate”. In: *Risk magazine* (2012).

[Car12b] Laurie Carver. “OCC faces VAR vetting questions over JP Morgan loss”. In: *Risk magazine* (2012).

[CCGIrt] Vincent P. Crawford, Miguel A. Costa-Gomes, and Nagore Iriberri. “Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications”. In: *Journal of Economic Literature* (forthcoming).

[CHH02] Murray Campbell, A. Joseph Hoane, Jr, and Feng hsiung Hsu. “Deep Blue”. In: *Artificial Intelligence* 134.1–2 (2002), pp. 57–83.

References IV

[CS04] Vincent Conitzer and Tuomas Sandholm. "Self-interested automated mechanism design and implications for optimal combinatorial auctions". In: *Proceedings of the 5th ACM conference on Electronic commerce*. EC '04. New York, NY, USA: ACM, 2004, pp. 132–141. DOI: 10.1145/988772.988793.

[DL01] John J. Donohue III and Steven D. Levitt. "The Impact of Legalized Abortion on Crime". In: *Quarterly Journal of Economics* 116.2 (2001), pp. 379–420.

[ES10] Larry G. Epstein and Martin Schneider. "Ambiguity and asset markets". In: *Annual Review of Financial Economics* 2 (2010), pp. 315–346.

[Fel74] Martin S. Feldstein. "Social Security, Induced Retirement, and Aggregate Capital Accumulation". In: *Journal of Political Economy* 82.5 (1974), pp. 905–926.

[FG05] Christopher L. Foote and Christopher F. Goetz. *Testing Economic Hypotheses with State-Level Data: A Comment on Donohue and Levitt*. working paper 05-15. FRB Boston, 2005.

References V

[GE09] Umberto Grandi and Ulle Endriss. "First-Order Logic Formalisation of Arrow's Theorem". In: *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI-2009)*. Ed. by X. He, J. Horty, and E. Pacuit. Lecture Notes in Artificial Intelligence 5834. Springer, 2009, pp. 133–146.

[GE11] Christian Geist and Ulle Endriss. "Automated search for impossibility theorems in social choice theory: ranking sets of objects". In: *Journal of Artificial Intelligence Research* 40 (2011), pp. 143–174.

[Gea01] John D. Geanakoplos. *Three brief proofs of Arrow's impossibility theorem*. Discussion Paper 1123RRR. New Haven: Cowles Foundation, 2001.

[Gea05] John D. Geanakoplos. "Three brief proofs of Arrow's impossibility theorem". In: *Economic Theory* 26.1 (2005), pp. 211–215.

[Gib73] Allan Gibbard. "Manipulation of voting schemes: a general result". In: *Econometrica* 41.4 (1973), pp. 587–601.

[Gon08] Georges Gonthier. "Formal proof – the four color theorem". In: *Notices of the AMS* 55.11 (2008), pp. 1382–1393.

References VI

[GS01] Itzhak Gilboa and David Schmeidler. *A Theory of Case-Based Decisions*. Cambridge University Press, 2001.

[Gun+01] John A. Gunnels et al. "FLAME: Formal Linear Algebra Methods Environment". In: *ACM Transactions on Mathematical Software* 27.4 (2001), pp. 422–455.

[Hal05] Thomas C. Hales. "A proof of the Kepler conjecture". In: *Annals of Mathematics* 162.3 (2005), pp. 1063–1185.

[Har06] John Harrison. *Floating-Point Verification using Theorem Proving*. Ed. by Marco Bernardo and Alessandro Cimatti. Bertinoro, Italy, 2006.

[Har12] John Harrison. "The HOL Light theory of Euclidean space". In: *Journal of Automated Reasoning* (2012). online first.

[HMT71] Léon Henkin, James Donald Monk, and Alfred Tarski. *Cylindric algebras, Part I*. Vol. 64. Studies in Logic. North Holland, 1971.

[Jor06] J. S. Jordan. "Power and legitimacy in pillage games". mimeo. 2006.

References VII

[Kle10] Paul Klemperer. "The product-mix auction: a new auction design for differentiated goods". In: *Journal of the European Economic Association* 8.2–3 (2010), pp. 526–536.

[KMP12] Michael Kohlhase, Bogdan A. Matican, and Corneliu C. Prodescu. "MathWebSearch 0.5 – Scaling an Open Formula Search Engine". In: *Intelligent Computer Mathematics*. Conferences on Intelligent Computer Mathematics (CICM) (Bremen, Germany, July 9–14, 2012). Ed. by Johan Jeuring et al. LNAI 7362. Springer Verlag, 2012. URL: http://kwarc.info/kohlhase/submit/aisc12-mws.pdf.

[KP84] Yakar Kannai and Bezalel Peleg. "A note on the extension of an order on a set to the power set". In: *Journal of Economic Theory* 32.1 (1984), pp. 172–175.

[KR09] Manfred Kerber and Colin Rowat. *Stable sets in three agent pillage games*. Working Paper 09-07. University of Birmingham, Department of Economics, 2009.

[Kre74] V. L. Kreps. "Bimatrix games with unique equilibrium points". In: *International Journal of Game Theory* 3 (1974), pp. 115–118.

References VIII

[KRW11] Manfred Kerber, Colin Rowat, and Wolfgang Windsteiger. "Using *Theorema* in the Formalization of Theoretical Economics". In: *Intelligent Computer Mathematics: 18th Symposium, Calculemus and 10th International Conference, MKM 2011*. Ed. by James H. Davenport et al. Lecture Notes in Artifical Intelligence 6824. Springer-Verlag, 2011, pp. 48–73.

[KT92] A. Kats and J.F. Thisse. "Unilaterally competitive games". In: *International Journal of Game Theory* 21.3 (1992), pp. 291–299.

[LL82] Dean R. Leimer and Selig D. Lesnoy. "Social Security and Private Saving: New Time-Series Evidence". In: *Journal of Political Economy* 90.3 (1982), pp. 606–629.

[LP10] Christian List and Ben Polak. "Introduction to judgment aggregation". In: *Journal of Economic Theory* 145.2 (2010), pp. 441–466.

[LR82] William F. Lucas and M. Rabie. "Games with no solutions and empty cores". In: *Mathematics of Operations Research* 7.4 (1982), pp. 491–500.

[Luc68a] William F. Lucas. "A game in partition function form with no solution". In: *SIAM Journal of Applied Mathematics* 16.3 (1968), pp. 582–585.

References IX

[Luc68b] William F. Lucas. “A game with no solution”. In: *Bulletin of the American Mathematical Society* 74.2 (1968), pp. 237–239.

[McC09] Bruce D. McCullough. “The accuracy of econometric software”. In: *Handbook of Computational Econometrics*. Ed. by David A. Belsley and Erricos John Kontoghiorghes. Wiley, 2009. Chap. 2, pp. 1–55.

[McC10] Bruce D. McCullough. “Econometric computing with R”. In: *Advances in social science research using R*. Ed. by Hrishikesh D. Vinod. Springer, 2010. Chap. 1, pp. 1–22.

[McC97] William McCune. “Solution of the Robbins problem”. In: *Journal of Automated Reasoning* 19.3 (1997), pp. 263–276.

[Min11] Yaron Minsky. “OCaml for the masses”. In: *ACM Queue* 9.9 (2011), pp. 44–53.

[MR90] Paul Milgrom and John Roberts. “Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities”. In: *Econometrica* 58.6 (1990), pp. 1255–1277.

References X

[Nip09] Tobias Nipkow. "Social choice theory in HOL: Arrow and Gibbard-Satterthwaite". In: *Journal of Automated Reasoning* 43.3 (2009), pp. 289–304.

[Rub98] Ariel Rubinstein. *Modeling bounded rationality*. Zeuthen lecture book series. MIT Press, 1998.

[Sat75] Mark Allen Satterthwaite. "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions". In: *Journal of Economic Theory* 10.2 (1975), pp. 187–217.

[SL09] Michael Schmidt and Hod Lipson. "Distilling Free-Form Natural Laws from Experimental Data". In: *Science* 324.5923 (2009), pp. 81–85.

[Suz00] Kotaro Suzumura. "Welfare economics beyond welfarist-consequentialism". In: *Japanese Economic Review* 51.1 (2000), pp. 1–32.

References XI

[Suz02] Kotaro Suzumura. "Introduction". In: *Handbook of Social Choice and Welfare*. Ed. by Kenneth J. Arrow, Amartya Sen, and Kotaro Suzumura. Vol. 1. Handbooks in Economics. Elsevier Science, 2002. Chap. 1, pp. 1–32.

[TL09] Pingzhong Tang and Fangzhen Lin. "Computer-aided proofs of Arrow's and other impossibility theorems". In: *Artificial Intelligence* 173.11 (2009), pp. 1041–1053.

[TL11a] Pingzhong Tang and Fangzhen Lin. "Discovering theorems in game theory: two-person games with unique pure Nash equilibrium payoffs". In: *Artificial Intelligence* 175.14–15 (2011), pp. 2010–2020.

[TL11b] Pingzhong Tang and Fangzhen Lin. "Two equivalence results for two-person strict games". In: *Games and Economic Behavior* 71.2 (2011), pp. 479–486.

[VLO06] René Vestergaard, Pierre Lescanne, and Hiroakira Ono. *The inductive and modal proof of Aumann's theorem on rationality*. technical report IS-RR-2006-009. Japan Advanced Institute of Science and Technology, 2006.

References XII

[Wie07] Freek Wiedijk. "Arrow's impossibility theorem". In: *Journal of Formalized Mathematics* 15.4 (2007), pp. 171–174.

[Wie09] Freek Wiedijk. "Formalizing Arrow's theorem". In: *Sādhanā* 34.1 (2009), pp. 193–220.

[Woo+09] Jim Woodcock et al. "Formal method: practice and experience". In: *ACM Computing Surveys* 41.4 (2009), pp. 1–40.

[ÅHW09] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. "Reasoning about coalitional games". In: *Artificial Intelligence* 173.1 (2009), pp. 45–79.

[ÅHW11] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. "On the logic of preference and judgment aggregation". In: *Autonomous Agents and Multi-Agent Systems* 22.1 (2011), pp. 4–30.

Glossary I

automated reasoning

automated theorem proving

calculus of inductive constructions [VLO06]

interactive theorem proving (proof assistant; human directed)

infix notation

logic

- first order (predicate)
  - many sorted
- second order
- higher order
- modal "the logic to reason about binary relations" [ÅHW11]
- propositional

machine learning

Glossary II

mathematical knowledge management

mechanized reasoning

model checker given a model, Γ, and a property, φ, does Γ have property φ (i.e. $\Gamma \vDash \varphi$)?

model theory

proof checker

quantifier elimination

SAT solver

semantic v syntactic

semantic web

set theory

simply typed λ-calculus HOL Light is built on top of this

situation calculus

skolemization

temporal calculus

Robbins' conjecture with two atoms, $\mathcal{E} = \{a, b\}$

X	Y	$\overline{X}$	$\overline{Y}$	$\overline{X} \vee Y$	$\overline{X} \vee \overline{Y}$	$\overline{\overline{X} \vee Y}$	$X \vee \overline{Y}$	HUN	ROB
0	0	1	1	1	1	1	1	0	0
0	a	1	b	1	1	b	b	0	0
0	b	1	a	1	1	a	a	0	0
0	1	1	0	1	1	0	0	0	0
a	0	b	1	b	1	b	1	a	a
a	a	b	b	1	b	b	1	a	a
a	b	b	a	b	1	0	a	a	a
a	1	b	0	1	b	0	a	a	a
b	0	a	1	a	1	a	1	b	b
b	a	a	b	a	1	0	b	b	b
b	b	a	a	1	a	a	1	b	b
b	1	a	0	1	a	0	b	b	b
1	0	0	1	0	1	0	1	1	1
1	a	0	b	a	b	0	1	1	1
1	b	0	a	b	a	0	1	1	1
1	1	0	0	1	0	0	1	1	1

◂ Robbins

Main software libraries

- *Mizar*
  - Mizar Mathematical Library (MML) has formalized 49,000 theorems
  - declarative proof mode resembles human mathematics: describes steps, rather than tactics (procedural)
  - outputs to the Journal of Formalized Mathematics
- HOL Light [Har12]
  - 9,724 named formal theorems (including trivial, e.g. $\pi > 0$)
  - built-in first order theorem prover, `MESON`
  - HOL Light Euclidean library developed out of *FlysPecK* contains Brouwer's fixed point theorem, Stone-Weierstrass, Tietze extension theorem, second mean value theorem for integrals, power series for real and complex transcendental functions, . . . Generally does not reach beyond $\mathbb{R}^n$ to arbitrary Banach, Hilbert spaces
  - lacks results in algebraic topology, differential forms, differential manifolds

Other encodings of UN, IIA and ND

First order logic [TL09] [q.v. GE09]

$$\forall a, b, s. [\forall x\, p(x, a, b, s)] \supset w(a, b, s) \quad (\mathtt{UN})$$

$$\forall a, b, s_1, s_2. [\forall x. p(x, a, b, s_1) \equiv p(x, a, b, s_2)] \supset [w(a, b, s_1) \equiv w(a, b, s_2)] \quad (\mathtt{IIA})$$

$$\neg\exists x \forall s, a, b. p(x, a, b, s) \equiv w(a, b, s) \quad (\mathtt{ND})$$

Higher order logic [Nip09]

$$\text{If } \forall i. Pia < Pib \text{ then } FPa < FPb \quad (\mathtt{UN})$$

$$\text{If } \forall i. Pia < Pib \leftrightarrow P'ia < P'ib \text{ then } FPa < FPb \leftrightarrow FP'a < FP'b \quad (\mathtt{IIA})$$

Modal logic [ÅHW11]

$$\Box\blacksquare((1 \wedge \cdots \wedge n) \to \sigma) \quad (\mathtt{UN})$$

$$\Box \bigwedge \blacksquare((o \wedge \sigma) \to \Box(o \to \sigma)) \quad (\mathtt{IIA})$$

$$\bigwedge_{i \in N} \Diamond\blacklozenge\neg(\sigma \leftrightarrow i) \quad (\mathtt{ND})$$

▸ Arrow by induction