

Algorithmic Game Theory: Artificial Agents Play and Learn, Producing Surprising and Instructive Results

Steven O. Kimbrough
kimbrough (à) wharton.upenn.edu

11 June 2008

Singapore Management University

[http://opim-sky.wharton.upenn.edu/~sok/sokpapers/
2009/SMU20080611/comprats-master-foils.pdf](http://opim-sky.wharton.upenn.edu/~sok/sokpapers/2009/SMU20080611/comprats-master-foils.pdf)

Case Examples

1. Spatialized Prisoner's Dilemma, page 27
 2. Q-learning in 2×2 games, page 32
 3. Monopoly Stories, page 36
 4. Cournot competition, page 40
 5. Bertrand competition, page 41
 6. Electricity Prices, page 43
 7. Two-sided matching, page 44
- Other models, page 45

Acknowledgements

I am most grateful for the contributions of many collaborators, especially Robert L. Axtell, James D. Laing, Fred Murphy, and David Harlan Wood. Of course, I claim exclusive rights to all errors contained herein.

Abstract

Scientific study of strategic interaction proceeds with four main methodologies. (i) Classical game theory has for the most part been conducted analytically, as a branch of applied mathematics. (ii) Behavioral game theory (and economics) generates and models data from laboratory experiments. (iii) Studies *in vivo* draw upon field or historical observations. (iv) Algorithmic studies model players as effective procedures and investigate the resulting collective behaviors. The theory of games and strategic interaction is enriched by the interaction of these perspectives.

This presentation focuses on results obtained through recent work in algorithmic studies of strategic interaction. When games are indefinitely repeated the predictions of classical game theory become much less

specific. Robust, more concrete predictions can be made if we assume that agents engage in simple, plausible learning regimes. We report and discuss the significance of results from standard 2×2 games, from oligopoly models involving 2 and more than 2 agents, and from bargaining and coalition formation models.

Informs Abstract

Title (100 of 100 characters): Algorithmic Game Theory: Artificial Agents Play, Learn and Produce Surprising, Instructive Results

Abstract (498 of 500 characters): We report results from recent work in algorithmic studies of strategic interaction. When games are indefinitely repeated the predictions of classical game theory become much less specific. Robust, more concrete predictions can be made if we assume that agents engage in simple, plausible learning regimes. We report and discuss the significance of results from standard 2×2 games, from oligopoly models involving 2 and more than 2 agents, and from bargaining and coalition formation models.

Greetings and Goals

This talk: about simple computational models that bear on fundamental research into rationality and decision making in strategic contexts.

Will spend some time to communicate the context and motivation, and ultimately the significance of this line of research. In doing so, I will (attempt to) keep it in mind that this is a computationally-oriented audience, rather than an economics-oriented one.

My message is very positive for computational investigations.

Context: Two kinds of decisions

- Parametric decisions. Decision analysis.
≈ “When other people’s decisions don’t matter.”
- Strategic decisions. Game theory.
≈ “When other people’s decisions do matter.”

Note: In both cases much conceptual development has been based on Rational Choice Theory (≈ von Neumann-Morgenstern utility theory, and SEU theory).

This presentation is about: Strategic decision making.

Methodological Starters

Observation 1:

Nearly all interesting decision problems are computationally complex.

Observation 2:

If a problem is computationally complex, then it must be attacked with heuristics.

(Note: (i) We can quibble, but the essential points remain. (ii) The points are valid for any form of high computational cost.)

Methodological Starters

Observation 3:

Analytic results for heuristics are always welcome, but are of limited availability.

Methodological Starters

Observation 4:

Progress on solving (better: handling) computationally difficult problems can be made by vigorous (and rigorous) empirical investigation of relevant heuristics, aided by insight-based intuition.

Methodological Starters: Comments

- Example work illustrating “vigorous (and rigorous) empirical investigation of relevant heuristics, aided by insight-based intuition” includes a series of papers by Kimbrough and Wood on their “feasible-infeasible 2-population GA” (FI-2Pop GA) for constrained optimization. See [KKLW08].
- Observations 1–4 are well accepted for constrained optimization problems/models, as well as other challenging *parametric decision problems* that are important for applications.

Methodological Starters: Comments

- Evolutionary computation and other forms of metaheuristics (simulated annealing, tabu search, scatter search, ant colony optimization, artificial immune systems, etc.) are in fact being vigorously investigated for these problems, are meeting with considerable success, and are securing considerable uptake in practice.
- In this talk I focus on *strategic* decision making, where the observations have historically not been honored (but should be and now are becoming accepted).

The problem? Finding good policies of play. How should one play?
Let's look at some examples.

Consider: How would you play this game?

| | Left | Right |
|------|------|-------|
| Up | 3 3 | 0 5 |
| Down | 5 0 | 1 1 |

If played once? If played 10 times with the same counter-player? If played 100 times? If played indefinitely?

This is an instance of a Prisoner's Dilemma game

| | Left | | Right | |
|------|---------|---------|-------|--|
| Up | 3 [P] 3 | 0 [P] 5 | | |
| Down | 5 [P] 0 | 1 [N] 1 | | |

P = Pareto optimal; N = Nash equilibrium (for the one-shot game).

Characteristic of Prisoner's Dilemma: separation of N and P.

Axelrod's tournaments and their results [Axe84] . . .

And why couldn't game theory settle the question

of how to play Repeated Prisoner's Dilemma?

- Solution concept: equilibrium (Nash or refinement)
- Folk Theorem: with (indefinite) repetition, nearly everything is in some equilibrium.

Indiscriminate prediction. Applies to *all* repeated games, not just PD.

Consider: How would you play this game?

| | Left | Right |
|------|------|-------|
| Up | 3 3 | 0 2 |
| Down | 2 0 | 1 1 |

If played once? If played 10 times with the same counter-player? If played 100 times? If played indefinitely?

Called a Stag Hunt game

Up, Left = “hunt stag”; Down, Right = “hunt hare”. Important conceptually. Note:

| | Left | | Right | |
|------|--------|---|-------|---|
| Up | 3 [NP] | 3 | 0 | 2 |
| Down | 2 | 0 | 1 [N] | 1 |

Two Nash equilibria (in pure strategies, plus a mixed equilibrium of $\langle \frac{1}{2}, \frac{1}{2} \rangle$) for the one-shot game. Drift of classical game theory: Down-Right, mutual hunting of hare.

Note vexing problems: (1) If we predict the outcome is an equilibrium, and there is more than one, how do the players possibly find the same equilibrium? (2) Proliferating equilibria...

Learnability, illustrated with Stag Hunt

- You play Stag Hunt repeatedly against a fixed counter-player. How hard can it be to learn how best to play?
- It can be impossible. Your counter-player might rely on information events that you do not have access to. Or might encode via its play a Turing machine and use it to generate future play. Or might change its policy of play from time to time. Or all of the above.
- We can hope for progress by examining heuristics and searching for robustness.

And what about other games and strategic decisions?

These are just simple cases, yet they are not adequately treated by current theory.

- Setting prices and quantities
- Choosing a partner
- Chess, checkers
- &c

From Binmore [Bin92, pages 50–1]

What is important here is that game theory does not pretend to tell you how to make judgments about the shortcomings [in terms of ideal rationality] of an opponent. **In making such judgments, you would be better advised to consult a psychologist than a game theorist.** Game theory is about what players will do when it is understood that both are rational in some [specific, very strong] sense. . . .

Does that mean that game theory is useless? Obviously I do not think so or I would not devote my time to it. It is however true that, unless there are good reasons for supposing that the people involved will behave rationally, game theory cannot realistically be used in a naive way to make predictions about what real people will

do. As a consequence, a player would often be unwise to use the strategy that a game theory book may label as “optimal” because this will usually only be optimal if *everyone* plans to play optimally. Of course, there are circumstances in which it *is* reasonable to work on the hypothesis that people will behave in a reasonably rational manner. Economics is somewhat shafily founded on the assumption that this will typically be the case in commercial and business transactions. However, it would be skating on very thin ice to use game theory for predictive purposes if none of the following criteria were satisfied:

- The game is simple.
- The players have played the game many times before¹ and hence have had much opportunity for trial-and-error learning.
- The incentives for playing well are adequate.

¹Against different opponents each time. If you are to play a particular game against the same opponent many times, one must model the repeated situation as a single “super-game”.

...However, even when all three criteria are satisfied, game-theoretic predictions can only realistically be applied with great caution.

Problems with classical game theory

Except for the two-person constant-sum game, it has long been recognized that there is no unique solution concept that can be regarded as *the* natural extension of individual rational behavior to multiperson decision making. This is one of the central problems of game theory. [Shu82, page 369]

Problems with classical game theory

Ironically, [classical] game theory is often hoisted on its own pétard: many of its most fundamental predictions—predictions that would have been too vague to test with any confidence in the pre-game-theoretic era—are *decisively and repeatedly disconfirmed*, in laboratory settings, with substantial agreement among experimenters, regardless of their theoretical priors. [Gin00, page xxiv]

Context: Ways of studying strategic decision making

- A priori. Classical game theory. Game theory as a branch of applied mathematics. Now also “algorithmic game theory” in the sense of finding computational means to compute difficult equilibria.
- In vitro. Behavioral game theory. Experiments in the lab. Empirical study of play.
- In vivo. “Games in the wild.” Observe and record.
- In silico. Evolutionary game theory. Epistemic game theory; computational heuristics. (i) Properties and behavior of strategies? (ii) of learning regimes? (Note: learning as randomized search. Fictitious play is not learning on this definition.)

From an interview with Robert Altman

Questioner:

Why do you hate Hollywood?

Altman:

I don't hate Hollywood. They make shoes. I make gloves. We're in different businesses.

Spatialized Prisoner's Dilemma, play C or play D

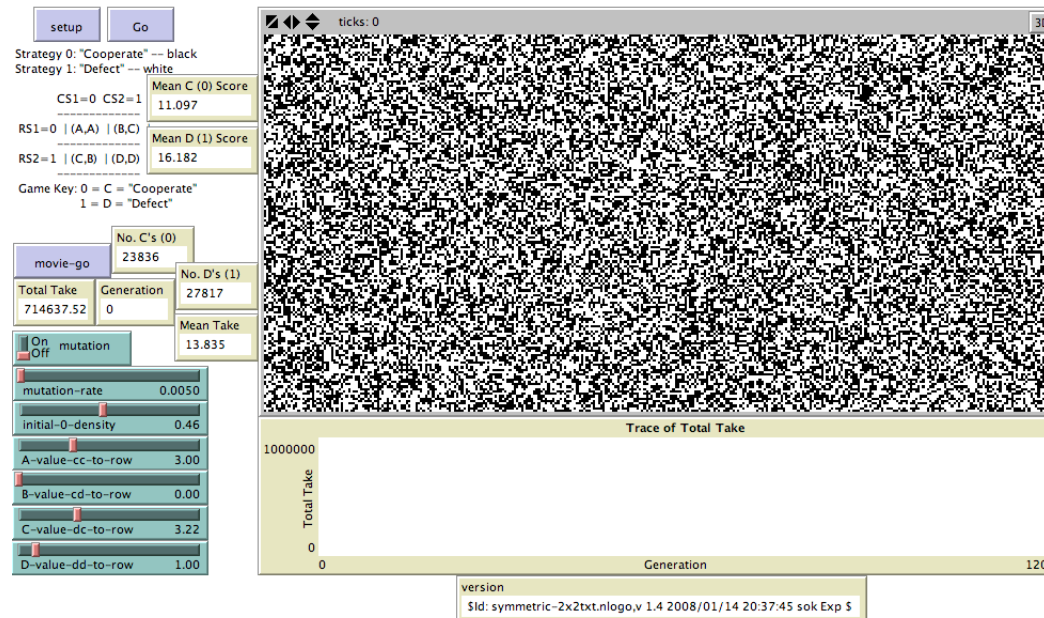


Figure 1: Prisoner's Dilemma initialized on the gridscape network.

Play your 8 neighbors; imitate the strategy of the most successful.

Cooperators devastated

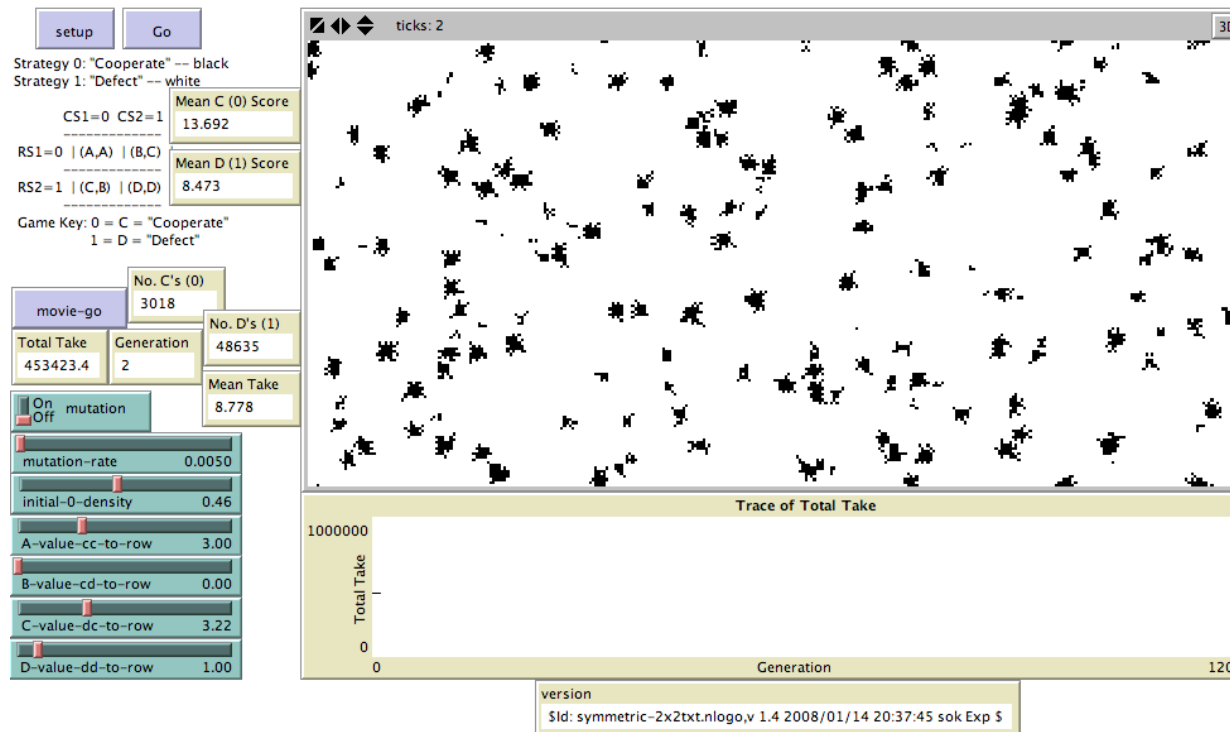


Figure 2: Prisoner's Dilemma generation 2: cooperators are decimated.

Surprise, surprise

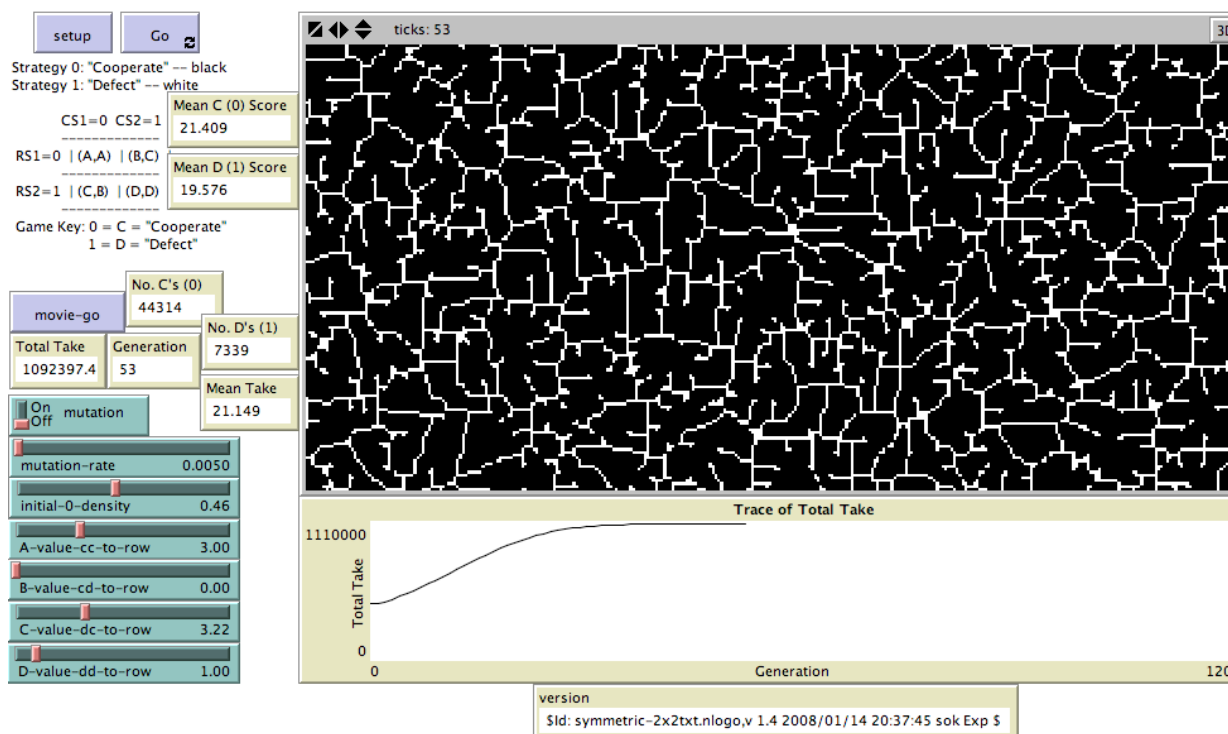


Figure 3: Prisoner's Dilemma generation 53: cooperators conquer.

And they stay that way

SOK's "invasion inequalities" for 2×2 games: In d dimensions, a hypercube of S_1 s can invade an abutting field of S_2 s if

$$A[2 \cdot 3^{d-1} - 1] + B3^{d-1} > \max\{D[2 \cdot 3^{d-1} - 1] + C3^{d-1}, D[3^d - 1]\} \quad (1)$$

Similarly, a hypercube of S_2 s can invade an abutting field of S_1 s if

$$D[2 \cdot 3^{d-1} - 1] + C3^{d-1} > \max\{A[2 \cdot 3^{d-1} - 1] + B3^{d-1}, A[3^d - 1]\} \quad (2)$$

| | S_1 | | S_2 | |
|-------|-------|-----|-------|-----|
| S_1 | A | A | B | C |
| S_2 | C | B | D | D |

$$A = 3, B = 0, C = 3.22, D = 1, d = 2$$

in our example, and $15 = 5A > \max\{5D + 3C, 8D\} = 14.66$.

Not a fluke. Important points.

- Algorithmic study of strategic interaction can yield surprising (and valid) results.
- Association between/among agents conduces cooperation.
- The regime of play matters greatly.

See [Kim08] for a brief presentation of the invasion inequalities and the Prisoner's Dilemma example.

Q-learning in 2×2 games

“Simple Reinforcement Learning Agents: Pareto Beats Nash in an Algorithmic Game Theory Study” [KL05]

<http://opim-sky.wharton.upenn.edu/~sok/sokpapers/2005/kimbrough-lu-pareto-beats-nash.pdf>

SSR: strategy selection regime. WE, wealth extracted is the proper measure of performance (ratio scale).

Backed off from Ideal Rationality and from Ideal Knowledge. We favor: LPS: learning in policy space.

Q-learning study. Finding: Pareto predicts better than Nash.

1. Alternative (or consideration) set.

In the 2×2 case, conditioning on the last play by the counter-player, $\mathcal{A} = \{[0, 0], [0, 1], [1, 0], [1, 1]\}$ for each player. Interpretation of $[x, y]$: if on the last play the counter-player played x , on this play I play y , where 1=C and 0=D.

2. Attractiveness estimation.

E.g., linear updating rule for A^i , $i \in \mathcal{A}$:

$$A_{t+1}^i = A_t^i + \alpha \{r_t^i - A_t^i\}$$

NewEstimate = CurrentEstimate + StepSize{reward - CurrentEstimate}

In the case of Q-learning, $Q'(s, a) = Q(s, a) + \alpha(r - Q(s, a))$

3. Choice/exploration policy.

ϵ -greedy. By default $\epsilon = 0.05$.

Figure 4: Summary of Elements in a Q-Learning Regime for 2×2 Games

Consider now play by two Q-learning agents of Indefinitely Repeated Prisoner's Dilemma. The stage game is repeated 100,000 times in each run. This counts as indefinite repetition because the agents have no, and can represent no, information concerning the length of play. We parameterize the stage game, as in Figure 5.

| | C | D |
|---|-------------------|----------------------|
| C | $(3,3)^*$ | $(0, 3+\delta)^*$ |
| D | $(3+\delta, 0)^*$ | $(\delta, \delta)\#$ |

Figure 5: A Parameterized PD Stage Game. $\#$ =Nash; $*$ =Pareto

| CC | CD | DC | DD | δ | Row's % CC |
|------|-----|-----|------|----------|------------|
| 9422 | 218 | 183 | 177 | 0.05 | 0.963 |
| 9036 | 399 | 388 | 150 | 0.5 | 0.963 |
| 5691 | 738 | 678 | 2693 | 1 | 0.931 |
| 3506 | 179 | 275 | 6040 | 1.25 | 0.972 |
| 1181 | 184 | 116 | 8519 | 1.5 | 0.930 |
| 2 | 98 | 103 | 9797 | 1.75 | 0.805 |
| 97 | 114 | 91 | 9698 | 2 | 0.735 |
| 0 | 100 | 92 | 9808 | 2.5 | 0.839 |
| 2 | 96 | 94 | 9808 | 2.95 | 0.986 |

Table 1: Summary of Results for Prisoner's Dilemma. ϵ -greedy action selection. Totals are for the last 100 rounds of 100 series of 10,000 plays.

Monopoly Stories

See <http://opim-sky.wharton.upenn.edu/~sok/age/nlogo/monopolyProbeAndAdjust.html>.

PROBE AND ADJUST. Random walks. A better account?

The classic account of monopoly offers much insight into the phenomena. Remarkably, it does this without appeal to any data other than the common observation that sellers seem to do better in, and seem to prefer, monopolistic markets than in seemingly competitive markets with many suppliers. The story, however, is hardly a complete account of monopoly. As one author notes with regard to the standard—or “classical”—theory of monopoly,

That's it. That's all there is to the standard theory of monopoly. We think of the monopoly being able to select the quantity it will supply or, equivalently, the price it will charge, and the resulting equilibrium price or quantity is read off the demand curve. But this simple classic theory is subject to a lot of questions, caveats, variations, and elaborations. [Kre90, page 301]

The standard monopoly story has pluses. It explains higher prices, etc. But it also has minuses.

- How does the monopolist know the demand curve? Is this realistic?
- What about changing demand?

PROBE AND ADJUST offers an alternative explanation that in many ways handles the “minuses” of the standard account.

Probe and Adjust

- KISS philosophy.
- PROBE AND ADJUST is a simple, plausible, accessible learning regime for adjusting a *continuous* parameter.
- How it works...
- Learning to choose among a discrete consideration set is also important and much more studied.

(In the interests of brevity, I'll mainly stick to PROBE AND ADJUST applications in this talk.)

Monopoly pricing/quantity setting is not strategic. Let's look at a closely-related setting that is: oligopoly.

Cournot competition

- Classic model of oligopoly. Here, firms offer quantities of a good and the market sets the price.
- The classic theory is undermotivated mathematically. Assumes “best response” behavior. This leads to the Cournot equilibrium, which lies between the monopoly quantity and the competitive quantity.
- See [KM07], “Learning to Collude Tacitly on Production Levels by Oligopolistic Agents”.
- Let’s look at <http://opim-sky.wharton.upenn.edu/~sok/age/nlogo/oligopolyPutQuantity.html>.

Bertrand competition

- Now the firms compete on price.
- Economics theory: collusion is impossible. Even with just two firms in the market they will compete away their profits.

Note the business literature on this: Don't do it!

- **See** <http://opim-sky.wharton.upenn.edu/~sok/age/nlogo/oligopolyBidPrice.html>

Management Implications?

- Executive pay? How to incentivize managers?
By comparison with competitors? By returns to the firm?
- Once again, standard theory comes up short.

Electricity Prices

- Firms bid a step function. The ISO (independent system operator) sums the bids horizontally and the market realizes the price. Not tractable analytically.
- Whereas extant mainstream literature tends to minimize the possibility of tacit collusion on prices, we find collusion all too easily established.
- **See** <http://opim-sky.wharton.upenn.edu/~sok/age/nlogo/ElectricityBidPrices.html>.

Two-sided matching

G-S, “triumph of game theory”

[AK08]

stability, fairness, social welfare

Other models

IDS (interdependent security) games; SPD (stochastic Prisoner's Dilemma). And generally games with stochastic payoffs. (With Howard Kunreuther's group.)

Marriage and negotiated surplus. (With Rob Kurzban.)

Small economies. (With Michael Weisberg.)

Conclusion

What have we learned, what can we learn, from algorithmic studies of strategic interaction? Much, including:

- Insight on a big question: Is a given market efficient? What are its operating characteristics? Can agents collude or do competitive conditions prevail?
- Insight on another big question: How should one play the game? How can high quality policies of play be discovered?
- Under what conditions can cooperation emerge? or be destroyed?

- Surprising behavior can emerge, even under rather simple conditions.
- Explaining behavior on the basis of procedures/rules accessible to agents is not only possible, it is preferable, even necessary.

Even more fundamentally, algorithmic studies teach us that we need to reconsider our concept(s) of rationality itself.

Conclusion: Rationality

The theory of games, indeed much of social science, has been built upon Rational Choice Theory, yet this theory has led to many unsatisfactory results. I am hardly the first to question it. Sen's critique, "Rational Fools" [Sen77], remains apt. Here is recent statement of his.

Rationality is interpreted here, broadly, as the discipline of subjecting one's choices—of actions as well as of objectives, values and priorities—to reasoned scrutiny. Rather than defining rationality in terms of some formulaic conditions that have been proposed in the literature (such as satisfying some prespecified axioms of "internal consistency of choice," or being in conformity with "intelligent pursuit of self-interest," or being some variant of maximizing behavior), rationality is seen here in much more general terms as the need to subject one's choices to the demands of reason. [Sen02, page 4]

Conclusion

The simple models and experiments illustrated here point towards a fundamental rethinking of the concept(s) of rationality in strategic contexts and indeed to a fundamental rethinking of how we investigate the social sciences.

Rationality not as optimization, but as a prudent tradeoff between exploration and exploitation.

Rationality that is accessible to us.²

²What some philosophers call an *internalist* view, in distinction to an *externalist* view. But that's another story.

References

- [AK08] Robert L. Axtell and Steven O. Kimbrough, *The high cost of stability in two-sided matching: How much social welfare should be sacrificed in the pursuit of stability?*, Proceedings of the 2008 World Congress on Social Simulation (WCSS-08), 2008, <http://mann.clermont.cemagref.fr/wcss/>.
- [Axe84] Robert Axelrod, *The evolution of cooperation*, Basic Books, Inc., New York, NY, 1984.
- [Bin92] Ken Binmore, *Fun and games: A text on game theory*, D.H. Heath and Company, Lexington, MA, 1992.
- [Gin00] Herbert Gintis, *Game theory evolving: A problem-centered*

introduction to modeling strategic interaction, Princeton University Press, Princeton, NJ, 2000.

- [KA06] Steven O. Kimbrough and Robert L. Axtell, *On concepts of rationality in games*, Society for Exact Philosophy, 2006.
- [Kim08] Steven O. Kimbrough, *On artificial intelligence, evolutionary computation and networks*, Proceedings of the Wharton-INSEAD Alliance Conference on Network-based Strategies and Competencies, 2008.
- [KKLW08] Steven Orla Kimbrough, Gary J. Koehler, Ming Lu, and David Harlan Wood, *On a feasible–infeasible two–population (FI-2Pop) genetic algorithm for constrained optimization: Distance tracing and no free lunch*, European Journal of

Operational Research **190** (2008), no. 2, 310–327, DOI information: 10.1016/j.ejor.2007.06.028.

- [KL05] Steven O. Kimbrough and Ming Lu, *Simple reinforcement learning agents: Pareto beats Nash in an algorithmic game theory study*, Information Systems and e-Business **3** (2005), no. 1, 1–19, <http://dx.doi.org/10.1007/s10257-003-0024-0>.
- [KM07] Steven O. Kimbrough and Frederic H. Murphy, *Learning to collude tacitly on production levels by oligopolistic agents*, working paper, University of Pennsylvania, Department of Operations and Information Management, September 2007.
- [Kre90] David M. Kreps, *A course in microeconomic theory*, Princeton University Press, Princeton, NJ, 1990.

- [Sen77] Amartya K. Sen, *Rational fools: A critique of the behavioural foundations of economic theory*, *Philosophy and Public Affairs* **6** (1977), 317–344.
- [Sen02] Amartya Sen, *Rationality and freedom*, ch. Introduction: Rationality and Freedom, pp. 3–64, Harvard University Press, Cambridge, MA, 2002.
- [Shu82] Martin Shubik, *Game theory in the social sciences*, The MIT Press, Cambridge, MA, 1982.

End Note

See also my `surprise-exam-foils.tex` and `gatech-20040305-foils.tex`. SEP paper/presentation La Jolla, spring 2006 (7?), [KA06].

Discuss (later): LPS, learning in policy space.

\$Id: comprats-master-foils.tex 350 2008-06-10 08:43:36Z sok \$