Constitutional Political Economy, 12, 81–84, 2001.

Reviews

H. PEYTON YOUNG (1998) *Individual Strategy and Social Structure: An Evolutionary Theory*. Princeton, NJ: Princeton University Press, pp. xiii, 189. $37.50.

Modern neoclassical economics is built on two pillars: (1) agents make optimal choices consistent with a completely specified maximand, and (2) the aggregate consequences of these individual choices are equilibria stable and unique enough to permit prediction. A glance at the profession's leading journals readily confirms the ascendancy of the maximization *cum* equilibrium method.

At the same time, few economists believe, least of all those most adept with the maximization *cum* equilibrium method, that real agents actually can and do make decisions with recourse to LaGrange and Hamilton, and that real markets quickly settle down into equilibria well-behaved enough for reliable prediction. And with good reason. For one thing, the experimental evidence is generally disconfirming. Theoretically, too, we know that cognitive resources are scarce, that nature is usually uncooperative in providing well-defined decision problems, and that in strategic settings multiple equilibria are ubiquitous.

What is more, modern economic method is ordinarily silent on social structures. Institutions such as laws, norms, and conventions, and organizations such as firms and governments are typically made exogenous, beyond analysis. This removes from economics political economy's great project of explaining spontaneous order: how Paris gets fed, or how the decentralized, interactive decisions of imperfectly informed agents come to be beneficially coordinated without collective human design.

None of this will be news to readers of this journal. What is news is that the mainstream of the profession, notably in game theory, is beginning, half a century on, to heed Herbert Simon's call for a more empirically-grounded economics. Peyton Young's monograph is an outstanding contribution to this new literature, not only because it offers a more realistic depiction of strategic decision making and its consequences, but also because, in so doing, it provides a compelling account of how social structures emerge and evolve.

Of the difficulties that beset game theory in particular, the problem of multiple equilibria is probably the most fundamental. John Nash shared a Nobel Prize for his dissertation's proof that in non-cooperative games an equilibrium exists. But he did not show how to choose among them when there were more than one, a task eventually taken up in the Nash refinements literature. For the most part, the refinements literature hewed closely to classical game theory's tacit requirement that a unique equilibrium must be determined by rational analysis *alone*—game theory prescribes what ideally rational players, given only a complete description of the game, must choose. Harsanyi and Selten (1988), a collaboration of the two theorists who shared the Nobel with Nash, best exemplifies this approach.

The Nash refinements program foundered, in part because it produced too many solution concepts—how is the agent to choose among multiple *theories* of equilibrium selection? Consider a coordination game, where two randomly paired traders can use one of two currencies, gold (G) or silver (S). If both play G, each gets a payoff of two; if both play S, each gets a payoff of one. If they fail to coordinate, each gets zero. The superhuman agent of classical game theory cannot decide what to do. But real people can; they coordinate by recourse to convention. They trade in gold (or silver), drive on the right (or left) side of the road, determine price by haggling (or by posting or by auction), use standard-form contracts, etc. Conventions coordinate expectations when deduction by itself is insufficient.

It's clear that conventions exist and that their coordinating function is economically valuable, so much has been known since Hume. Thomas Schelling's (1960) contribution was to see necessity behind their virtue. *Homo sapiens* is smarter than *homo economicus* precisely because *homo sapiens* is less rational. Boundedly rational agents, unlike their idealized cousins, have incentives to look for extrasomatic help.

But how is it that conventions become conventional? Game theory long neglected the question, because of its tacit commitment to solution via deduction alone: a uniquely rational solution obviates the need for conventions, so players who observe conventions cannot be ideally rational, and, conversely, should it prove rational to follow a convention, then the claim of a uniquely rational solution must be false. Not surprisingly, alternative approaches were developed far afield, in theoretical biology.

An evolutionarily stable strategy (ESS) is a pattern of behavior that cannot be successfully invaded by a group of mutants, because mutants would obtain lower payoffs. Clearly, all ESS are Nash, but not all Nash equilibria are ESS. In the currency coordination game, let x be the proportion of Gold users, and $(1 - x)$ that of Silver users. A Silver user has an expected payoff of $(1-x)$ and a Gold user has an expected payoff of $2x$, so a Silver strategy is "fitter" when $(1 - x) > 2x$, or when less than 1/3 of the population uses Gold. The case where exactly 1/3 of the population uses Gold is unstable, a tipping point. A shock that increases the number of Gold users from 1/3 will push all traders towards Gold, and a shock that increases Silver users from 2/3 will push all traders towards Silver; both are ESS. (This all-or-none outcome arises only because fitness is increasing with the proportion using the strategy. This routinely is not the case, as in the Hawk and Dove game, for example). Once the Gold standard is established, deviants who play Silver cannot successfully invade, and a Silver standard is likewise impervious to Gold users. This asymptotic stability is characteristic of an ESS, as is the possibility of "lock-in," the stability of Pareto-inferior equilibria such as the Silver standard

Peyton Young was among the first economists to attempt an evolutionary approach. In it, maximization and equilibrium are replaced with adaptive rationality and dynamic evolution, respectively. "Adaptive" means that players respond to incentives, but not perfectly. Players are not assumed to have common knowledge (i.e., mutual knowledge of infinite degree) of the game structure, other players' utility functions and rationality. Instead they have incomplete information, bounded memory and simple models of how others are likely to behave.

"Evolution" means that the process of convergence is dynamic. The process itself is modeled; players are not assumed to instantaneously deduce ("select") an equilibrium. "Neoclassical economics describes what the world looks like once the dust has settled," says Young. "We are interested in how the dust goes about settling." Adaptively rational players player grope towards an equilibrium, pushed by a changing environment as much as pulled by their own deductive prowess. In addition, players are drawn from a population of potential player types, and the probability of their interaction with others depends upon physical (or other social) proximity. Finally, the choice process is buffeted by a stream of persistent small shocks, analogous to the role of biological mutation.

This last aspect distinguishes Young's main analytical concept—stochastic stability—from ESS. Young argues that perturbing shocks (e.g., errors (ε) made by players) are likely to be steady, rather than occasional as assumed in asymptotic models, so that a sufficient accumulation of adverse shocks can dislodge the prevailing standard, tipping from Gold to Silver, say. In truly dynamic systems, even a small of amount of noise can generate long-run outcomes quite different from those in an equivalent deterministic system. A stochastically stable outcome is one that is robust given shocks of frequency ε.

Game theorists will note that Young's low-rationality, evolutionary approach obtains many of the outcomes of classical game theory, and does so with vastly less heroic assumptions. Harsanyi and Selten's risk-dominant equilibrium turns out to be stochastically stable in coordination games, for example. Most conspicuously, Young's theory suggests that more efficient conventions will predominate over time. The Gold standard rules most of the time, with relatively short punctuated interruptions of Silver, a significant departure from asymptotically stable concepts, which allow inefficient conventions to reign. Intuitively, efficient conventions are more resistant because the series of shocks required to push the Silver users' proportion to 2/3 is less likely than the series required to push Gold users to 1/3.

Young's last two chapters extend the theory to bargaining and contracts. Bargaining conventions are common—real estate agents get six percent (paid by the seller), the lawyer's contingency fee is one-third, and experimental subjects split the pie fifty-fifty. How do bargaining conventions become established? As before, the answer is by the accumulation of precedent, a feedback process where past experience shapes current expectations. Here again, the theory recovers something very close to the Nash bargaining solution, and does so without the requirement of common knowledge. And (paraphrasing) if anything is common knowledge, it's that common knowledge isn't common.

Young here makes good on two pioneering ideas. He offers a rigorous account of Schelling's insight that coordination results not in spite of low rationality, but because of it. And he also makes good on Oskar Morgenstern's notion that rules are not merely the prerequisites of games, but can also be regarded as the product of games.

Young's monograph, which has its origins in five lectures given to the Institute of Advanced Study at Hebrew University, delivers the theoretical goods. But Herbert Simon's critique is ultimately an empirical one. Simon's point is less about the paucity of realistic theories than about the unwillingness to meaningfully confront our many models empirically. Young has a better story, but we now want to know how it fares. In particular, there is the matter, which Young scrupulously flags, of the rate of institutional change. How often, for example, does an inefficient convention emerge and how long does it last? Is the Windows operating system entrenched for the foreseeable future, or will Justice break up Microsoft just as Linux becomes predominant?

Theoretically, the rate of change depends on where the rationality dial is set, the shock rate, the size of the population, and the extent of interaction. But for empirical relevance, we require a time scale, a metric for translating game time into historical time. An interaction is the unit of account, but how long *is* 30,000 interactions? In New York it's a minute, in Antarctica, it's an eon. It is unfair, of course, to ask that this excellent monograph provide the answers, but these are the answers that economics should want.

Thomas C. Leonard
Department of Economics
Princeton University
Princeton, NJ 08544
United States

Reference

Harsanyi, John, and Selten, Reinhard (1988) *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Schelling, Thomas (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.