Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 10564–10567, September 1999

Perspective

Stochastic game theory: For playing games, not just for doing theory

Jacob K. Goeree[†] *and Charles A. Holt*

Department of Economics, Rouss Hall, University of Virginia, Charlottesville, VA 22903

Recent theoretical advances have dramatically increased the relevance of game theory for predicting human behavior in interactive situations. By relaxing the classical assumptions of perfect rationality and perfect foresight, we obtain much improved explanations of initial decisions, dynamic patterns of learning and adjustment, and equilibrium steady-state distributions.

About 50 years ago, John Nash walked into the office of the Chair of the Princeton Mathematics Department with an existence proof for N-person games that was soon published in these *Proceedings* (1). John von Neumann dismissively remarked "that's trivial, you know, that's just a fixed point theorem" (ref. 2, p. 94). But word of Nash's work spread quickly, and with the Nash equilibrium as its centerpiece, game theory has now gained the central role first envisioned by von Neumann and Morgenstern (3). Game theory is rapidly becoming a general theory of social science, with extensive applications in economics, psychology, political science, law, and biology.

There is, however, widespread criticism of theories based on the "rational choice" assumptions of perfect decision making (no errors) and perfect foresight (no surprises). This skepticism is reinforced by evidence from laboratory experiments with financially motivated subjects. Nash participated in such experiments as a subject and later designed similar experiments of his own but lost whatever confidence he had in game theory when he saw how poorly it predicted human behavior (2). And Reinhard Selten, who shared the 1995 Economics Nobel Prize with Nash and Harsanyi, remarked that "game theory is for proving theorems, not for playing games" (personal communication). This paper describes new developments in game theory that relax the classical assumptions of perfect rationality and foresight. These approaches to introspection (before play), learning (from previous plays), and equilibrium (after a large number of plays) provide complementary perspectives for explaining actual behavior in a wide variety of games.

Coordination and Social Dilemma Games

The models summarized here have been influenced strongly by data from experiments that show disturbing differences between game-theoretic predictions and behavior of human subjects (4–7). Here, we consider a social dilemma game for which Nash's theory predicts a unique equilibrium that is "bad" for all concerned and a coordination game in which any common effort level is an equilibrium. That is, the Nash equilibrium makes no clear prediction.

The social dilemma is based on a story in which two travelers lose luggage with identical contents. The airline promises to pay any claim in an acceptable range as long as the claims are equal. If not, the higher claimant is assumed to have lied and both are reimbursed at the lower claim, with a penalty, R, being taken from the high claimant and given to the low claimant. A Nash equilibrium in this context is a pair of claims that survives an "announcement test": if each person writes in their claim and then announces it, neither should want to reconsider. Note that, when $R > 1$, each traveler has an incentive to "undercut" any common claim. For example, suppose the range of acceptable claims is from 80 to 200; then, a common claim of 200 yields 200 for both, but a deviation by one person to 199 raises that person's payoff to $199 + R$. The paradoxical Nash equilibrium outcome of this "traveler's dilemma" is for both travelers to claim 80, the lowest possible amount (8). Simple intuition suggests that claims are likely to be much higher when the penalty parameter, R, is low (8), but that claims may approach the Nash prediction when R is high. Fig. 1 shows the frequency of actual decisions in the final five rounds of an experiment with randomly matched student subjects for $R = 10$ (red bars), $R = 25$ (yellow bars), and $R = 50$ (blue bars) (9). The task for theory is to explain these intuitive treatment effects, which contradict the Nash prediction of 80, independent of R.

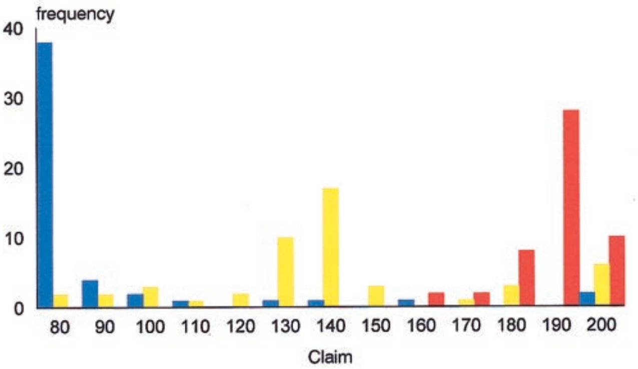

FIG. 1. Experiment based on the traveler's dilemma game, using randomly matched student subjects who made claim decisions independently in a sequence of 10 periods. Earnings ranged from $24 to $44 and were paid in private, immediately after the experiment. The frequency of actual decisions for the final five periods is indicated by the blue bars for $R = 50$, by the yellow bars for $R = 25$, and by the red bars for $R = 10$. With $R = 50$, the average claim was quite close to the Nash prediction of 80, but, with $R = 10$, the average claim started high (at ≈180) and moved away from the Nash prediction, ending up at 186 in the final five periods. The observed average claim varies with the penalty/reward parameter R in an intuitive manner (a higher R results in lower claims), in sharp contrast with the Nash prediction of 80, independent of R.

The second game is similar, with payoffs again determined by the minimum decision. Players choose "effort levels," and both have to perform a costly task to raise the joint production level. A player's payoff is the minimum of the two efforts minus the cost of the player's own effort: $\pi_i = \min\{x_1, x_2\} - cx_i$, where x_i is player i's effort level and $c < 1$ is a cost parameter. Consider any common effort level: A unilateral increase is

[†] To whom reprint requests should be addressed. E-mail: jg2n@virginia.edu.

costly but does not affect the minimum, and a unilateral decrease will reduce the minimum by more than the cost reduction because $c < 1$. Hence, any common effort survives the Nash announcement test.

A high effort choice is riskier when the effort cost is high, which suggests that actual behavior might be sensitive to the effort cost. Fig. 2 shows the time-sequences of average effort choices for three groups of 10 randomly matched subjects who made effort choices in the range from 110 to 170 cents (J.K.G. and C.A.H., unpublished work). There is an upward pattern for the low-cost treatment (thin blue lines for $c = 0.25$) and an essentially symmetric downward adjustment for the high-cost treatment (thin green lines for $c = 0.75$). The thick color-coded lines show average efforts for each treatment. The Nash equilibrium or any standard variant of it does not predict the strong treatment effect, which is consistent with simple intuition about the effects of effort costs. Another interesting result is that subjects may "get stuck" in a low-effort equilibrium (10, 11).

For both games, the most salient feature of observed behavior is not predicted by classical game theory. To understand these results, note that players are usually unsure about what others will do and that there may be randomness in the way individuals process and act on information. Even relatively small payoff asymmetries and small amounts of noise can have large "snowball" effects in an interactive, strategic game, and this intuition is a key element of the models presented below.

Evolutionary Dynamics

A natural approach to explaining divergent adjustment patterns like those in Fig. 2 is to develop models of evolution and learning (12–18). The idea behind evolution in this context is not based on the notion of reproductive fitness but, rather, that people tend to adopt successful strategies. Evolutionary models typically add stochastic elements that are reminiscent of biological mutation. Consider a population of players that are characterized by their decision, $x(t)$, at time t, which earns an expected payoff of $\pi^e\ (x(t),t)$ where the expectation is taken with respect to the current population distribution of decisions, $F\ (x,t)$. The evolutionary assumption is that players' decisions tend to move toward the direction of higher expected payoffs.

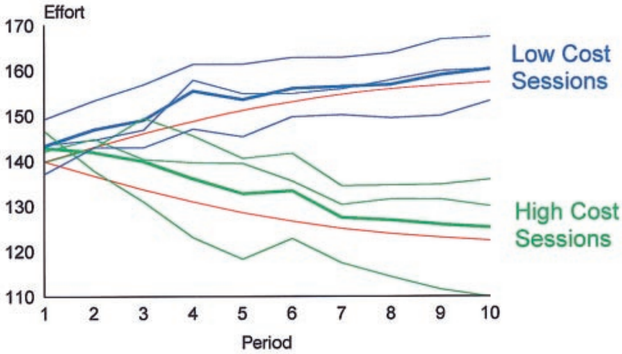

FIG. 2. Experiment based on the coordination game, using randomly matched student subjects who made effort decisions independently in a sequence of 10 periods. Average effort choices are given by the thin green lines for the low-cost sessions (with $c = 0.25$) and by the thin blue lines for the high-cost sessions (with $c = 0.75$). The thick lines show average efforts for both treatments. As simple economic intuition would suggest, higher effort costs result in lower efforts, an effect that is not predicted by the Nash equilibrium. Note that, in one of the high-cost sessions, effort choices gravitate toward the lowest possible effort choice of 110, the equilibrium that is worst for all concerned. The thin red lines give the average effort choices predicted by an evolutionary adjustment model in which players tend to move in the direction of higher payoffs but may make mistakes in doing so.

Hence, a player's decision will increase if the expected payoff function is increasing at $x(t)$, and vice versa. So the decisions evolve over time, with the rate of change proportional to the slope of expected payoff, $\pi^{e\prime}$, plus a stochastic Brownian motion term: $dx = \pi^{e\prime}(x(t),t)dt + \sigma\, dw(t)$, where σ determines the relative importance of the random shock $dw(t)$. It can be shown that this individual adjustment process translates into a differential equation for the population distribution:

$$\frac{\partial F(x,t)}{\partial t} = -\pi^{e\prime}(x,t)f(x,t) + \mu f'(x,t), \qquad [1]$$

where $\mu = \sigma^2/2$ is a noise or error parameter and f, f' represents the population density and its slope. This is the famous Fokker-Planck equation from statistical physics, which has been derived in this interactive context (S. P. Anderson, J.K.G., and C.A.H., unpublished work). The evolutionary process in Eq. **1** is a nonlinear partial differential equation that can be solved by using numerical techniques. The red lines in Fig. 2 show the resulting trajectories of average efforts for the two treatments of the coordination game experiment. This model explains both the strong treatment effect and the general qualitative features of the time path of the average effort choice.

Learning Dynamics

Evolutionary models are sometimes criticized on the grounds that they ignore the cognitive abilities of human subjects to learn and adapt. The learning model that is closest to the evolutionary approach is "reinforcement learning" based on the psychological insight that successful strategies will be reinforced and used more frequently. Whereas the previous section describes the evolution of decisions in a "large" population, reinforcement learning pertains to changes in decision probabilities of a single decision maker. Reinforcement models are formalized by giving each decision an initial weight and then adding the payoff actually obtained for a decision chosen to the weight for that decision (7, 19). The model is stochastic in the sense that the probability that a decision is taken is the ratio of its own weight to the sum of all weights. Simulations show that reinforcement learning models can explain key features of data from many economics experiments (19).

Alternatively, learning can be modeled in terms of beliefs about others' decisions. For example, suppose beliefs are characterized by a weight for each possible value of the other player's decision and that the subjective probability associated with each decision is its weight divided by the sum of all weights. Weights are updated over time by adding 1 to the weight of the decision that is actually observed while the other weights remain unaltered. The resulting subjective probabilities determine expected payoffs, which in turn determine choices. To allow for some randomness in responses to payoff differences, psychologists have proposed probabilistic choice rules that have the desirable property that decisions with higher expected payoffs are more likely to be chosen, although not necessarily with probability 1. In particular, the "logit" probabilistic choice rule specifies that the probability of selecting a particular decision, i, is proportional to an exponential function of its expected payoff, $\pi^e(i)$:

$$Pr(i) = \frac{\exp(\pi^e(i)/\mu)}{\sum_{j=1}^{N} \exp\ (\pi^e(j)/\mu)}, \quad i = 1, \ldots, N. \qquad [2]$$

As μ goes to infinity, all choice probabilities are equal, regardless of payoff differences, but small payoff differences will have large effects when μ goes to 0. Fig. 3 shows simulated

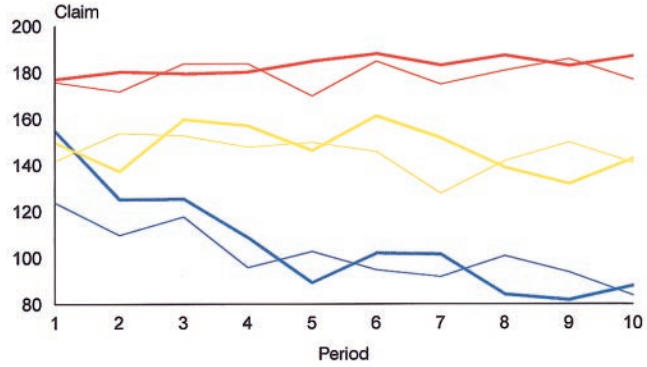

FIG. 3. Patterns of adjustment for the traveler's dilemma game. The thick lines show the average claims of human subjects in sessions with a penalty/reward parameter of $R = 50$ (blue line), $R = 25$ (yellow line), and $R = 10$ (red line). The color-coded thin lines represent simulated average claims for the different treatments. These simulations are based on a Bayesian learning model, in which players use observations of their opponents' past play to update their beliefs about what will happen next. Adding logit errors to such a naive learning model results in predictions that conform nicely with the actual adjustment patterns observed in the laboratory.

decisions with noisy learning, together with the observed choices of human subjects, for the three treatments of the traveler's dilemma game. Simulations of this type also have been used to explain adjustment patterns in other experiments (20, 21).

Logit Equilibrium

In a steady state, the distribution function in Eq. **1** stops evolving; that is, $dF/dt = 0$. Setting the right side of Eq. **1** to 0, dropping the time arguments, and rearranging yields a differential equation in the equilibrium density: $f'(x)/f(x) = \pi^{e\prime}(x)/\mu$. This equation can be integrated to yield $f(x) = k \exp(\pi^{e\prime}(x)/\mu)$, which is the continuous analogue of Eq. **2**. Thus, a gradient-based evolutionary model with Brownian motion produces the logit model in a steady state. Likewise, when simulated behavior in learning models with logit errors settles down, the steady state distributions of decisions are consistent with the logit rule in Eq. **2**.

In equilibrium, Eq. **2** must be interpreted carefully because the choice density determines the expected payoffs, which in turn determine the choice density. In other words, the logit equilibrium is a fixed point: The "belief" density that goes into the expected payoff function on the right side of Eq. **2** must match the "choice" density that comes out of the logit rule on the left side. It has been proven that such an equilibrium exists for all games with a finite number of decisions (22, 23). This elegant proof is based on a fixed-point theorem, like Nash's half-page proof in these *Proceedings* (1).

The equilibrium conditions in Eq. **2** can be solved numerically by using an estimated error parameter. Fig. 4 shows the logit equilibrium densities for the three treatments of the traveler's dilemma experiment. The theoretical densities pick up the general location of the data frequencies in Fig. 1, with the colors used to match predictions and data for each treatment. There are discrepancies, but the logit equilibrium predictions are superior to the Nash prediction of 80 for all treatments. The logit equilibrium has been successfully applied to explain behavior in a variety of other economic environments (refs. 22–25; R. D. McKelvey and T. R. Palfrey, personal communication).

A Model of Iterated Noisy Introspection

Many political contests, legal disputes, and special auctions are played only once, in which case there is no opportunity to learn, adjust strategies, and reach an equilibrium. One way to proceed in these situations is to use general intuition and introspection about what the other player(s) might do, what they think others might do, and so on. Edgar Allen Poe mentions this type of iterated introspection in an account of a police inspector trying to decide where to search for the "purloined letter." Economists have long thought about such iterated expectations and have noted that the resulting "infinite regression" often leads to a Nash equilibrium and hence does not provide a new approach to behavior in one-shot games.

Naturally, our approach is to introduce noise into the introspective process, with the intuition that players' own decisions involve less noise than their perceptions of others' decisions, which in turn involves less noise than other players' perceptions of others' decisions, and so on. Our model is based on the probabilistic choice mapping in Eq. **2**, which we will express compactly as $p = \phi_\mu(q)$, where q represents the vector of belief probabilities that determine expected payoffs, and p is the vector of probabilities that is determined by the probabilistic choice rule ϕ_μ. We assume that the error parameter associated with each higher level of iterated introspection is $\tau > 1$ times the error parameter associated with the lower level. For instance, $p = \phi_\mu\ (\phi_{\tau\mu}(q))$ represents a player's noisy (μ) response to the other player's noisy ($\tau\mu$) response to beliefs q. The "telescope" parameter τ determines how fast the error rate blows up with further iterations. We are interested in the choice probabilities as the number of iterations goes to infinity:

$$p = \lim_{n\to\infty} \phi_\mu(\phi_{\tau\mu}(\cdot\ \ \cdot\ \ \cdot\ \ \phi_{\tau^n\mu}(q))). \qquad \mathbf{[3]}$$

This limit is well defined for $\tau > 1$, and because ϕ_∞ maps the whole probability simplex to a single point, the limit is independent of the initial belief vector q. It has been shown (J.K.G. and C.A.H., unpublished work) that Eq. **3** provides a good explanation of (nonequilibrium) play in many types of one-shot games; see ref. 26 for alternative approaches (C. M. Capra, personal communication).

The logit equilibrium arises as a limit case of this two-parameter introspective model. Recall that a logit equilibrium is a fixed point of ϕ_μ: that is, a vector p^* that satisfies $p^* = \phi_\mu(p^*)$. For $\tau = 1$, a fixed point of ϕ_μ is also a fixed point of Eq. **3**. So, if the introspective model converges for $\tau = 1$, the result is a logit equilibrium. If $\tau > 1$, Eq. **3** determines beliefs

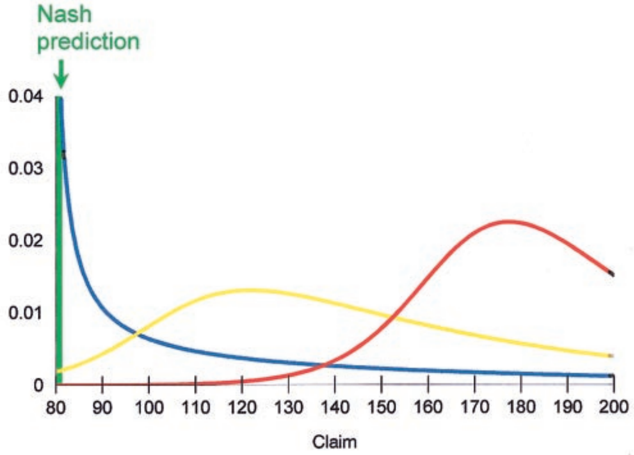

FIG. 4. Logit equilibrium densities for the traveler's dilemma game with a penalty/reward parameter of $R = 50$ (blue line), $R = 25$ (yellow line), and $R = 10$ (red line). The logit equilibrium is capable of reproducing the most salient feature of the human data: that is, the intuitive effect of the penalty/reward parameter on average claims. Note that the logit equilibrium predictions are roughly consistent with the actual human data shown in Fig. 1 and are far superior to the Nash prediction of 80 (green bar), which is independent of R.

p that will generally not match the iterated layers of introspective beliefs on the right side. In this sense, the introspective model generalizes the logit equilibrium by relaxing the assumption of perfect consistency between actions and beliefs, just as the logit equilibrium generalizes Nash by relaxing the assumption of perfectly rational decision making.

Conclusion

This paper describes three complementary modifications of classical game theory. The models of introspection, learning/evolution, and equilibrium contain the common stochastic elements that represent errors or unobserved preference shocks. Like the "three friends" of classical Chinese gardening (pine, prunus, and bamboo), these approaches fit together nicely, each with a different purpose. Models of iterated noisy introspection are used to explain beliefs and choices in games played only once, where surprises are to be expected, and beliefs are not likely to be consistent with choices. With repetition, beliefs and decisions can be revised through learning or evolution. Choice distributions will tend to stabilize when there are no more surprises in the aggregate, and the resulting steady state constitutes a noisy equilibrium.

These theoretical perspectives have allowed us to predict initial play, adjustment patterns, and final tendencies in laboratory experiments. There are discrepancies, but the overall pattern of results is surprisingly coherent, especially considering that we are using human subjects in interactive situations. The resulting models have the empirical content that makes them relevant for playing games, not just for doing theory.

We thank Melayne McInnes for a helpful suggestion. This research was funded in part by the National Science Foundation (Grants SBR-9617784 and SBR-9818683).

1. Nash, J. (1950) *Proc. Natl. Acad. Sci. USA* **36,** 48–49.
2. Nasar, S. (1998) *A Beautiful Mind* (Simon and Schuster, New York).
3. von Neumann, J. & Morgenstern, O. (1944) *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ).
4. Ochs, J. (1994) *Games Econ. Behav.* **10,** 202–217.
5. McKelvey, R. D. & Palfrey, T. R. (1992) *Econometrica* **60,** 803–836.
6. Beard, T. R. & Beil, R. O. (1994) *Management Sci.* **40,** 252–262.
7. Roth, A. E. & Erev, I. (1995) *Games Econ. Behav.* **8,** 164–212.
8. Basu, K. (1994) *Am. Econ. Rev.* **84,** 391–395.
9. Capra, C. M., Goeree, J. K., Gomez, R. & Holt, C. A. (1999) *Am. Econ. Rev.*, in press.
10. Van Huyck, J. B., Battalio, R. C. & Beil, R. O. (1990) *Am. Econ. Rev.* **80,** 234–248.
11. Cooper, R., DeJong, D. V., Forsythe, R. & Ross, T. W. (1992) *Q. J. Econ.* **107,** 739–771.
12. Foster, D. & Young, P. (1990) *Theor. Popul. Biol.* **38,** 219–232.
13. Fudenberg, D. & Harris, C. (1992) *J. Econ. Theor.* **57**(2), 420–441.
14. Kandori, M., George, M. & Rob, R. (1993) *Econometrica* **61,** 29–56.
15. Binmore, K., Samuelson, L. & Vaughan, R. (1995) *Games Econ. Behav.* **11,** 1–35.
16. Crawford, V. P. (1991) *Games Econ. Behav.* **3,** 25–59
17. Crawford, V. P. (1991) *Econometrica* **63,** 103–144.
18. Young, P. (1993) *Econometrica* **61,** 57–84.
19. Erev, I. & Roth, A. E. (1998) *Am. Econ. Rev.* **88,** 848–881.
20. Cooper, D. J., Garvin, S. & Kagel, J. H. (1994) *Rand J. Econ.* **106,** 662–683.
21. Camerer, C. & Ho, T. H. (1999) *Econometrica*, in press.
22. McKelvey, R. D. & Palfrey, T. R. (1995) *Games Econ. Behav.* **10,** 6–38.
23. McKelvey, R. D. & Palfrey, T. R. (1998) *Exp. Econ.* **1,** 1–41.
24. Anderson, S. P., Goeree, J. K. & Holt, C. A. (1998) *J. Political Econ.* **106,** 828–853.
25. Anderson, S. P., Goeree, J. K. & Holt, C. A. (1998) *J. Public Econ.* **70,** 297–323.

36. Harsanyi, J. C. & Selten, R. (1988) *A General Theory of Equilibrium Selection in Games* (MIT Press, Cambridge, MA)