Journal of Economic Theory **94**, 241–261 (2000)
doi:10.1006/jeth.2000.2701, available online at http://www.idealibrary.com on IDEAL®

NOTES, COMMENTS, AND LETTERS TO THE EDITOR

Economists' Models of Learning[1]

Joel Sobel

Department of Economics, University of California, San Diego, La Jolla, California 92093-0508
jsobel@ucsd.edu

Received July 10, 1998; final version received March 10, 2000

Economic theorists have intensively studied learning in games and decisions over the last decade. This essay puts some of the work in perspective and offers opinions about what still needs to be learned. *Journal of Economic Literature* Classification Numbers: C73, D83.

1. INTRODUCTION

Can imitation be the result of rational choice? Can people learn enough to make useful decisions by observing the choices and experiences of others? Will society be too slow to adopt useful technologies because of limited information? If there are multiple equilibria in a strategic model, perhaps reflecting low and high levels of economic activity, can the process by which agents approach equilibrium be used to predict the outcome? Can these forces be influenced by policy? How can economic institutions be designed so that people can learn more effectively? Can these institutions be designed to perform well even in the presence of incomplete learning?

Questions like these provide motivation for economists to study learning models. Agents in these models begin with a limited understanding of the problems that they must solve. Experience improves their decisions. Death and a changing environment worsen them. By discussing a variety of models, I hope to describe how learning has been modeled. I focus on identifying circumstances under which learning leads to optimal decisions. I do not

[1] This essay began as a collaboration with John Conlisk and could not have been written without his help. I thank Jonathon Bendor, Andreas Blume, Antonio Cabrales, Vincent Crawford, Yong-Gwan Kim, John McMillan, Garey Ramey, Akos Rona-Tas, Suzanne Scotchmer, Max Stinchcombe, and Joel Watson for comments and encouragement and NSF and CASBS for financial support.

0022-0531/00 $35.00

provide answers to the questions above. I may describe ideas that will ultimately provide answers.

I classify the learning setting on the basis of three characteristics: the strategic environment; the way in which agents collect information; and the degree of rationality of the agent. In Section 2, I discuss models in which there is a single agent. The category includes decision-theoretic models. In Section 3, I consider models in which there are many agents, but an individual agent cares about the existence and actions of other agents only because they provide information about the environment. The category includes models of observational and social learning. In Section 4, I discuss game-theoretical models in which the actions of one agent have a direct influence on the payoffs of the other agents.

Agents can collect information passively or actively; the collection process can be free or costly. Passive information collection takes place as an outcome of an adaptive process. Agents are unable to influence the quantity or quality of the information that they obtain. Learning is active when agents' choices determine the flow of information. An agent might do research on a product before deciding whether to buy it, or might make a decision in one period because it could lead to information that is of future importance. Active learning will be costly if it takes resources to acquire or process information. In the models that I discuss, the cost of learning is not a direct cost associated with purchasing information. Instead it is a cost associated with knowingly making a suboptimal decision in one period in order to obtain information that will improve future decision making.

Any discussion of learning must begin by assuming that agents need to learn something. Consequently, agents in learning models must fail to be fully informed or fully rational. Most of the models that I discuss explain the need for agents to learn because, initially, they lack information. Models of learning based on limited information neglect other possible reasons for learning. If agents have limited ability to solve constrained optimization problems, then the learning process could describe how they look for optima. If agents have limited ability to collect information, then the learning process could describe how they look for information. Biasing the discussion to models of bounded information reflects a not necessarily healthy bias in literature that I discuss.

A general rule in the papers that I discuss is that the more complicated the strategic environment, the simpler the information collection technology and the simpler the behavior followed by the agents. The decision-theoretic models in Section 2 assume that agents follow active learning rules. The agents have an accurate model of their environment and make optimal decisions within their model. In Section 3, when agents learn from the actions of other agents, information flow is passive and costless. In some models agents optimize. In other models agents follow behavioral rules of

thumb. When agents must learn in fully strategic environments, as they do in Section 4, agents determine their actions using a simple model of their opponents' behavior. Most learning in these strategic models is passive and free.

The combination of assumptions may be more motivated by theoretical aesthetics than descriptive power. When it is possible to solve a model without assuming that agents follow rules of thumb, theorists prefer to do so. I hope to suggest circumstances in which studying rules of behavior lead to more instructive, and possibly richer, models than the theoretical models based on narrow notions of rationality.

This essay is short. My knowledge is limited. I omit all but the most cursory reference to studies of learning from other disciplines.[2] Useful overviews of most of the work that I discuss exist. I hope that by referring you to these surveys, you will excuse me for not discussing some of the ideas and articles that they reference.[3]

2. INDIVIDUAL LEARNING

One strand of the learning literature concentrates on behavior after binary experiments. The prototypical example of this literature is the two-armed bandit problem, which was brought to attention to economists by Rothschild [53].[4] An agent must choose between two options each period. The job can be done badly or well. The choices differ in their quality, which is the probability they get the job done well. The agent does not know which choice is better; she does have a prior distribution that summarizes her prior information. Each day she makes a decision and observes whether the job is done badly or well. The agent's payoff is the discounted

[2] This footnote pays lip service to approaches to learning from other disciplines. Perhaps philosophers were the first to write about learning. Peirce's [50] ideas on scientific discovery and induction, written one hundred years ago, seem modern. Churchland [14] gives an overview of philosophical theories of learning and consciousness from the viewpoint of contemporary cognitive science. Psychologists have traditionally studied learning in animals (including humans) in controlled settings. Some recent work by economists recalls mathematical models of learning proposed by psychologists in the 1950s (Estes [20]). Connectionist models (neural networks, genetic algorithms, and classifier systems) provide models of learning that are attractive both because they appear to shed light on how the brain really works and because they provide algorithms that appear to solve problems efficiently in theory and in practice. Holland, Holyoak, Nisbett, and Thagard [35], a computer scientist, two psychologists, and a philosopher, propose a framework for studying learning. Hutchins [36], an anthropologist, discusses learning in social environments in a way that has much to offer economists interested in organizational behavior.

[3] A long interval between the preparation of this paper and its publication has led to a less excusable failure to discuss the most recent contributions to the literature.

[4] Berry and Fristedt [6] gives a systematic treatment of bandit problems.

sum of payments associated with the quality of the daily output. The quality of the job depends only on the decision, not on the experience or the actions of the agent. The agent only observes how her choice performed (not, for example, how the other choice would have performed). Information is valuable because it will enable the agent to pick the choice that is more likely to behave well. Information is costly because the only way in which the agent can find out how well a choice works is to make the choice. Provided that the agent discounts future payoffs and the qualities of the choices (probability of a good outcome) are independent and continuously distributed, the optimal policy involves (with probability one) switching choices finitely many times. Eventually, the agent will stick with one choice. Some of the time this choice will be the inferior one. The result follows because experimentation is costly. An agent who discounts would rather stick with a decision that may be wrong than pay the cost needed to find the correct decision.

In statistical learning models, in which information arrives in each period independent of the actions of the decision maker, an agent eventually obtains all available information and her actions are optimal given this information. If information arrives effortlessly, at a constant rate, full learning is inevitable in a stationary environment. Why should the models of rational learning in economics be any different? The potential to obtain new observations is constant over time in economic models, but individuals typically must pay a price (in terms of foregone utility) to take advantage of this information. An agent will not necessarily learn the optimal decision when the cost of acquiring additional information exceeds the benefit.

The qualitative predictions of the bandit problem may not describe actual behavior. Experimental subjects, both men and mice, tend not to settle on a single choice in a two-armed bandit problem. Ask an agent to repeatedly choose between two options one of which gives a reward with probability p and the other with probability q. Theory suggests that the optimizing agent should experiment for a while, and then repeatedly use the option that appears to provide the greatest probability of reward. Instead, subjects switch from option to option, playing the superior option with a higher probability. This qualitative outcome could be explained by naive learning models of psychologists (Estes [20]), in which choice is stochastic and the agent increases the probability that she plays an action that leads to a favorable outcome.[5] Failure to lock into an action could be the result of agents having bounded recall. It also could be because the agents misunderstand the problem that they are asked to solve (Simon [57]). An attractive hypothesis is that the experimental environment is

[5] Arthur [2] describes a related learning process that gives rise to a similar qualitative behavior.

artificial, and that experimental subjects are behaving in ways that lead to good (or acceptable) outcomes in natural settings.

The bandit problem assumes that the environment is stationary. Eventually the weight of experience will make it unprofitable to gather new information and the agent will lock into a particular strategy. This type of behavior is almost surely wrong in changing environments. To see this point, add to the example above a tiny bit of nonstationarity. To be specific, assume that there is a constant, positive probability that the probability that one program works well is redrawn from the original distribution. In this environment if it is optimal to try this program at first, then it will be optimal to try it infinitely often. If the agent ignores the variable program long enough, whatever information she gained from the original experimentation decays; it will be optimal to try it again. Provided that the probability that the environment changes is small the agent will periodically experience a sufficiently bad sequence of draws from the variable that she will return to the safe arm infinitely often as well. The lock in result of the bandit problem is quite special. In natural, nonstationary environments agents will experiment regularly.

Aghion, Bolton, Harris, and Jullien [1] provide a detailed analysis of the rational learning through experimentation model. The paper identifies situations in which the agent asymptotically learns to make the correct decision under general conditions on the action spaces, information spaces, and state spaces. In the model, nature first chooses the state. The agent does not learn the state. She makes decisions in all subsequent periods. After the agent makes a decision, she learns her payoff. The central question is whether, in the limit, the agent learns enough to make the optimal response to the true state. Complete learning requires strong assumptions. Aghion, Bolton, Harris, and Jullien [1] provide situations in which learning is complete. If the agent does not discount and experimentation is essentially free, then the agent will learn to make the optimal decision. If the agent knows that the payoff function is analytic or that it is smooth and quasi-concave, then local (and hence inexpensive) experimentation is sufficient to discover whether global improvement is possible. The authors point out that the restrictive assumptions needed for these results suggest that complete learning is not a likely outcome of optimal search. Even more interesting is the likely brittleness of the optimal search behavior. Agents would do badly if they behaved as if their payoff function were analytic even if it need not be.[6] How agents should act if they fully

[6] If agents were to act as if their payoff function were quadratic, then three local observations would be enough to determine the entire payoff function. So an experimenter might make three inexpensive experiments, compute the quadratic function consistent with the observations, and then jump to the global optimum. Such a procedure would work badly for non-quadratic payoff functions.

understand the decision problem and have unlimited computation power is likely to give little understanding of how a boundedly rational agent would and should behave.

3. SOCIAL LEARNING

The literature on social learning assumes that there are many agents. There is no strategic interaction between the agents; the actions of one agent do not directly influence the payoffs of another. The agents are informationally linked, so that the actions and payoffs of one agent provide information about the state of the world. Actions do not change the state of the environment. Agents do not choose how much information they receive. The model specifies what agents observe and when they observe it.

A simple subset of these models studies observational learning.[7] Agents make only one decision. A countable set of agents makes decisions sequentially. Each agent has access to information about the past decisions; an agent may have private information as well. Each agent's payoff depends only on the underlying state of the world and her decision; it does not depend directly on the decisions of other agents. Incomplete learning is a feature of these models, just as it was in the bandit problem described above. The literature has emphasized the possibility of herding, in which agents ignore their own private information, choosing instead to take the same action as earlier agents. I will begin the discussion with an informal story followed by a simple model that gives rise to herding.

Here is a simple formal model. There is a sequence of agents indexed by *i*. Agent *i* must make a binary decision in period *i*. Interpret the decision as whether to make an investment. Agent *i* observes the decisions of each of the agents that precedes her. In addition, each agent can observe a signal, which is the realization of a binary-valued random variable. Assume that the signals, s_i, are independently distributed and each signal is equally likely to take on the values 1 or -1. An agent receives $\sum \delta^{t-1} s_t$ if she makes the investment and $-\sum \delta^{t-1} s_t$ otherwise. Assume that $\delta \in (0, 1)$. Hence the test favors investing if the test result is positive. The true value of the investment is given by a weighted average of all of the test results. In this model, the first agent invests if and only if her signal is positive. All other agents do what the first agent did. Because of discounting, the second agent values the information of the first agent more than her own information. Hence she does what the first agent did. Future agents understand that only the first action conveyed information. They too can do no better than to follow the lead of the first agent. The collective decision could be

[7] Gale [30] provides an overview of rational observational learning models.

dramatically wrong. Plainly the model is special. I hope to describe some of the lessons that have and have not been learned from the study of more general models.

In the example agents may make a bad decision. Conceivably all but the first agent has private information suggesting that investing is a bad idea, yet each agent decides to invest. How sensitive is this conclusion to the special features of the example? Discounting (or imposing a cost of making an observation) plays only a small role.[8] It is not important that information is binary, but it is important that the set of available actions is discrete. If the set of choices is rich enough, one would expect an agent's optimizing action to reveal private information. If future agents process information in past actions effectively, then efficient actions result. So lumpiness of information is needed for inefficient herding. It is implausible that an action choice contains enough information to signal all private information (or if it did, that agents would be astute enough to take advantage of the signal). Lumpiness appears to be a sensible assumption.

Smith and Sørensen [60] show that inefficient herding could arise unless it is possible for agents to obtain a powerful enough private signal to counterbalance any observation of past behavior that they might obtain. This result illustrates a kind of fragility that has been pointed out by Bikhchandani, Hirshleifer and Welch [7]. Periodically an agent will receive a strong enough signal to reverse the information content of the actions of past individuals. When preferences are heterogeneous, it is natural to expect periodic reversals of fashion. These reversals also arise in models with homogeneous preferences provided that agents may obtain strong enough signals to reverse the trend. Everyone invests, say, until one agent's test result is so negative that she, and subsequent agents, switch. Smith and Sørensen [60] show that the tendency to switch away from the best choice goes to zero asymptotically when private signals can be unboundedly strong. Precisely when complete learning is guaranteed, convergence is guaranteed to be slow. Social learning may lead to the correct choice asymptotically, but important inefficiencies remain as one approaches the limit.

The individual and the observational learning literature share a common conclusion. Both identify situations in which after a finite number of periods, agents repeatedly make the same choice; the choice may be incorrect. The models have an important difference. Since agents make only one decision, no one gains by experimentation in the observational learning models. The decision maker is informed of past payoffs in much of the

[8] Without it (and assuming that there are finitely many agents so that the sum converges), as soon as there exists an n such that $s_{2n-1} = s_{2n}$ all subsequent agents ignore their private information.

learning literature. Smith and Sørensen [59] make a formal analogy between the observational learning model and a model of an impatient experimenter.

More general models of social learning include the possibility that different agents act at the same time or that agents make repeated actions, while maintaining the assumption that individuals do not control the amount of information that they receive. Models of word-of-mouth transmission of information parallel the observational learning literature. Banerjee and Fudenberg [4] analyze a model in which agents make one binary decision. Their payoff depends upon the information obtained from people who acted in the past. In each period agents obtain an unbiased sample of the population. The sample contains information about how many agents made each decision in the previous period. It also contains information about the payoffs associated with each decision. Banerjee and Fudenberg [4] prove that the system must converge to the efficient outcome provided that players receive information from at least two agents from the previous generation.[9] Banerjee and Fudenberg [4] use an informativeness condition that plays the same role of Smith and Sørensen's [60] assumption on the strength of individual signals. The assumption guarantees that with positive probability the information an agent gathers about payoffs will determine her action choice.

In Banerjee and Fudenberg's [4] model, agents do not choose the number of past observations that they receive; by assumption, this quantity is fixed. Otherwise, agents are fully rational. There have also been treatments of word-of-mouth learning with boundedly rational agents. Ellison and Fudenberg [19] study the asymptotic behavior in word-of-mouth learning models in which an agent takes the action A if, in her sample, all agents are using A or if A has a higher average payoff than B. Hence these agents neglect the information contained in the number of people in their sample who chose A over B and concentrate on the experience that these people had. While this behavioral rule typically responds to the experience of agents rather than the popularity of choices, it is sensitive to the choices of the population: An agent cannot change unless her sample contains at least one individual who chose the other action. Ellison and Fudenberg [19] identify conditions under which this behavior leads to efficient learning. Efficiency is not guaranteed, however. To get an intuition for this finding, imagine that each sample contains some agents who chose A and some agents who chose B. If A is only slightly better than B, then the average performance of B is likely to beat that of A in many of these samples; this pressure could prevent the fraction of agents who choose A from converging to one.

[9] Obtaining information from at least two agents guarantees that agents will sometimes choose an action based on the previous agents' payoff rather than only their action.

Ellison and Fudenberg [18] study a word-of-mouth model in which agents base decisions solely on the payoff information contained in their sample. They show that agents need not converge to the correct action, but do learn the correct action if their decision rule places an appropriate weight on the market shares of the competing choices. There will be situations where this information, which could be thought of as the share of the market held by the two alternatives, is relatively easy to infer. Hence it is sensible to consider situations in which adaptive consumers respond instead to market shares rather than to payoff information. Agents who do not fully use the information in the market shares of the two options will not necessarily converge to the correct decision. For this reason, Smallwood and Conlisk's [58] article provides an interesting contrast to Ellison and Fudenberg [18]. In Smallwood and Conlisk [58], payoff information comes in a simple form: choices lead to either a success or a failure. Choices differ in their probability of generating a successful outcome. Agents make repeat decisions based solely on market share. While the opportunity to make repeat decisions creates the possibility that agents may experiment, the decision-makers follow mechanical rules. Whatever information they acquire comes from their passive observation of market shares. If agents only change their action after a failure, and then choose a brand with probability equal to its market share, then asymptotically consumers choose only the best brands. Smallwood and Conlisk [58] demonstrate that this efficiency result is not robust to small changes in the decision rules. If agents are too sensitive to market share (so that the most popular choice attracts a share of consumers in excess of its market share) an inefficient choice that is originally popular could capture the market. The popularity of the item may attract new customers faster than its low quality loses them. If agents are relatively insensitive to market share, then different choices, having different qualities, maintain positive market shares in the limit. This model applies to a version of the technology adoption problem described above. Assume that agents have no private information, but contemplate changing word-processing programs whenever their program fails. In the limit agents will all make the better decision if, whenever an agent's choice fails, the agent makes each choice with a probability equal to that choice's current market share.

Ellison and Fudenberg [18] and Smallwood and Conlisk's [58] work demonstrates that limits on the rationality of agents changes the qualitative results of the models. Both pairs of authors suggest plausible ways in which agents may not fully solve the optimization problem that they face. A population of agents may choose a variety of different behavioral rules. Would asymptotic learning be more likely if some agents concentrated on the representation in sample and others on payoffs? Are some behavioral rules more likely to survive than others are? Do any of the behavioral rules

conform to the types of heuristics identified in experiments? Under what conditions would we expect agents to invest most of their time gathering opinions of others, which they evaluate with simple decision rules, rather than optimally processing the information in their samples? I do not know. If (in a binary-choice model) the better choice has the higher payoff most of the time (this depends upon the distribution of the idiosyncratic noise), then whenever all of the population uses average performance to make a decision, the better decision will have the greater market share. Hence market share is informative. If it were easy to observe market share, one would expect a positive fraction of the economy to use market share information to make their decisions. This observation is at the heart of Ellison and Fudenberg's [18] finding that behavioral rules that use both performance and market share are more likely to converge to the correct choice than rules that ignore market share information. It suggests that two types of agent (one type basing decisions on market share information, the other using payoff information) could co-exist and even improve learning performance.

These models study the implications of simple, plausible rules, but do not provide insight into where these rules come from. The process by which agents decide where to eat probably differs in substantial and relevant ways from the way in which new technologies are adopted. Existing models give little insight about how to describe the differences. While evidence from experiments has convinced theorists that it is useful to incorporate boundedly rational agents in their analyses, analytical convenience and introspection are the main justifications for the assumptions about behavior used in adaptive models. To the extent that experiments provide evidence of systematic deviation from rationality, models should more carefully reflect what has been learned from experiments.

The models of observational learning generate stark and illustrative predictions. Authors have used these models to provide theories to explain a range of phenomena from bank runs to dining choices of both humans and ants.[10] The models explain why large groups of academic economists write papers about informational cascades. The papers based on optimizing behavior neatly summarize the observation that individuals may neglect their own information and choose instead to follow the crowd. It is not clear that the optimizing models are more instructive than simple models of imitation that would lead to the same type of predictions, however. Indeed, a qualitative conclusion that one can take from the work is that social learning by sophisticated agents may be the same as the behavior of simple imitators.

[10] See Banerjee [3], Bikhchandani, Hirshleifer and Welch [7], and Kirman [41] for these applications.

Several aspects of the observational learning problem are important intuitively, but play no role in the formal model. The model does not provide any information about whether the decision problem is easy or hard. Efficient learning results may depend on agents having finer powers of inference (their updating is more complicated) or greater abilities to solve optimization problems (since they must select a value for a continuous variable rather than a binary one), but these notions are informal.

The possibility of incomplete learning in these models is not surprising. Individuals have information that is relevant to future decision makers, but they can communicate only through their actions. Yet people neglect the impact that their information has on future decision making. So even in models in which asymptotic efficiency can be guaranteed, the rate of convergence is much slower than what one would expect in a social optimum.[11]

4. LEARNING AND EVOLUTION IN GAMES

Learning models played an important role in earlier days of game theory. In the fifties, Brown [11] proposed fictitious play as a mechanical method to compute equilibria of zero-sum games. In fictitious play agents keep track of the empirical distribution of their opponents' strategy choices and respond optimally to that distribution. Relatively simple arguments demonstrate that under this process the empirical distribution converges to the distribution of maxmin points in two-player, zero-sum games (Robinson [52]). It was also known to converge to equilibria in generic, two-by-two games (Miyasawa [46]). In the early sixties, Shapley [56] provided a nonconvergence example using a two-player, nonzero-sum game in which each player had three strategies. For the most part game theorists ignored dynamics for the next twenty years.

There has been a renewed interest in dynamic models of game behavior recently. One motivation for the study is a frustration with the equilibrium methods that dominated the game theory produced and consumed during most of the seventies and eighties. The applied papers started with an economic situation and described it using a complicated game. Several canonical models involving bargaining or signaling had multiple equilibria. An industry arose that developed equilibrium-selection arguments to deal with these situations. Alas, the selections were not theoretically compelling; they imposed enormous requirements on agents' ability to reason (and enormous faith that other agents followed similar reasoning paths); they

[11] Vives [61] proves this type of result. Gale [30] emphasizes the possibility that optimal learning may be slow.

were difficult to work with; and the predictions that they generated were not always consistent with intuition, common sense, or experimental evidence. In place of relatively general models of rationality that provided equilibrium concepts for broad classes of games, game theorists have substituted extraordinarily simple models of individual choice. In place of the question: What would rational players do in this situation? Came the question: Would non-rational players of a particular type eventually do things that we ascribe to rational players (avoid strongly dominated strategies, play equilibrium strategies)? For the most part this work provides results that warn against using equilibrium models in applications. The sufficient conditions for convergence to equilibrium are strong.

I give an overview of some recent developments in this section. First, I discuss evolutionary models in which individual choice and learning play a secondary role. Asymptotic results depend upon the tendency of relatively successful strategies to grow. Second, I discuss learning models based on fictitious play. Some of these models permit individuals to learn through active, and costly, experimentation. Finally, I consider models in which agents have a Bayesian prior on their opponents' strategies.[12]

Evolutionary game models are close to models of social learning described in the previous section. While these models focus on strategic situations, the players are decidedly non-strategic. Agents change strategies infrequently. When they change, they usually adopt a strategy based upon its recent success. With small probability they "mutate" to a random strategy. Agents do not actively gather information. Their strategic choices do determine the payoffs to other strategies, and hence the future path of play. Identifying the differences between evolutionary models and social learning models may suggest modifications of the evolutionary models that make them more plausible starting points for economic analysis. Both classes of model assume that agents do not change their strategy until an opportunity arises. This opportunity is stochastic. In social learning models, however, the probability of changing strategies may depend in a systematic way on the strategy being used.

Evolutionary models have been used to make selections in general games with multiple, but nonstrict, equilibria. For games in which there is a unique efficient payoff, some adaptive learning creates a tendency for the system to drift away from any inefficient outcome. When the population reaches an efficient outcome it tends to stay there.[13]

[12] More detailed treatments of these ideas are available. See, for example, Börgers [10], Fudenberg and Levine [28], Kandori [39], or Marimon [62].

[13] These results seem suited to repeated games (Binmore and Samuelson [8], Fudenberg and Maskin [29], and Robson [52]) and games with communication (Nöldeke and Samuelson [49], Wärneryd [62], and Blume, Kim, Sobel [9]).

Kandori, Mailath, and Rob [40] and Foster and Young [21] study strategic, evolutionary models in which there is perpetual randomness. Imagine a finite population playing a two-by-two game with two strict equilibria. One adaptive process would have each player adjust its strategy by playing a best response to the previous population distribution. Under this procedure both strict equilibria would be locally stable. Now add to the adaptive process a probability of a random change in strategy. If many members of the population simultaneously change their strategies, then it is possible for the entire population to jump from one equilibrium to the other. When the random change has a small probability and the population is large, the probability of moving to another equilibrium is small. These articles demonstrate that the population spends almost all of its time near a particular equilibrium -loosely speaking, the equilibrium that it is easier to jump to, when there is a large population and little randomness. These models do select an equilibrium. The selection, however, depends strongly on arbitrary features of the model. For example, Bergin and Lipman [5] identify the importance of the assumption that the probability of random change does not depend on where the system is. Fudenberg and Harris [23] point out that results also depend on the fact that the discrete-time dynamic focuses attention on the length, rather than the depth, of the basins of attraction.[14] Furthermore, Ellison [17] has pointed out that the long-run predictions only are relevant for cockroaches, as all other life forms will have long been extinct before the system reaches its limit. He proposes models of local interaction that provide speedier adjustment (but not always the same results). The slow convergence results have the same flavor as the results from the observational learning literature (Smith and Sørensen [60]). The sharp predictions of these models come with another cost. Individual behavior is narrowly and mechanically defined. The connection between learning by goal oriented, albeit boundedly rational agents, and the dynamics studied in these papers is largely a hopeful analogy.

Selection arguments based on evolutionary models are subject to criticism that parallels the criticisms of refinements. Selection arguments based on refinements rely on agents that are too smart to be possible; selection arguments based on evolution rely on agents that are too dumb to survive against agents with even limited rationality. Refinements are sensitive to arbitrary details of the assumptions regarding rationality. Evolution to arbitrary details of the evolutionary process. Ultimately neither approach is theoretically compelling.

Learning models may provide weak long-run predictions about behavior in games. Players who form beliefs about their environment and then

[14] Models of mutations incorporating some of the considerations found in social learning models may make it possible to evaluate the force of these critiques.

choose strategies that are optimal responses to their beliefs, will not play strictly dominated strategies. If beliefs also are largely formed on the basis of experience, such that past experience that is not repeated is eventually given little weight, not only will players avoid dominated strategies, but they will not place weight on their opponents using dominated strategies. In this way one is able to conclude that the limits of these dynamics must be rationalizable strategies.[15] These arguments apply directly to fictitious play and to other adaptive models (Gul [34] and Milgrom and Roberts [45]). The arguments provide a foundation for predictions in a limited class of strategic environments.[16] These dynamics have a curious mix of rationality and simplicity. On one hand, agents use a false, backward-looking procedure to make forecasts about other players' behavior. On the other hand, these agents select an optimal response to these forecasts. Hence the models assume that learning about the environment is so difficult that agents make no attempt to learn other than to assume that a population statistic is adequate. The models also assume that learning how to solve an optimization problem is so simple that an agent will not settle for less than a best response. Furthermore, agents do not really learn how to figure out what their opponents are doing. Rather, the rule that they use for making forecasts turns out to be correct whenever opponents' behavior stabilizes. Other dynamics, in which agents do not optimize subject to beliefs, need not lead to even the minimal rationality restrictions in the limit as fictitious play and its relatives.[17]

Some of the literature on learning can be seen as an effort to revive the fictitious play model, or at least to identify more precisely when fictitious play converges to equilibrium.[18] The research program of Fudenberg and

[15] In two-player games rationalizable strategies are precisely those that survive iterated deletion of strongly dominated strategies.

[16] Milgrom and Roberts [44] provides results for supermodular and dominance solvable games.

[17] The replicator dynamic of evolutionary biology selects strategies that perform well, but do not necessarily respond optimally, to strategies played in the previous period. Dekel and Scotchmer [16] construct an example that demonstrates, in fact, that the replicator dynamic does not necessarily avoid dominated strategies. While the example depends on the discrete nature of the adjustment process (Samuelson and Zhang [54] and Cabrales and Sobel [13]) it strongly suggests that agents who are unable to compute best responses to beliefs need not converge to undominated strategies.

[18] The work of Foster and Young [22] and Monderer and Shapley [47] provide a deeper understanding of when fictitious play will converge. These papers extend the class of games that have the fictitious play property (that players acting according to the fictitious play algorithm actually converge to a Nash equilibrium) to dominance solvable games and games with identical payoffs. Fudenberg and Levine [27] show that smoothed versions of fictitious play have a strong consistency property. Players who adjust their beliefs according to the empirical distribution of play and then respond optimally to their beliefs choose acts that, asymptotically, are approximate best responses to the empirical distribution of play. It appears that the rate of convergence is slow.

Kreps ([24] and [25]) is dedicated to finding conditions on learning models that guarantee that local stability of Nash equilibria (particularly mixed strategy equilibrium, for the local stability of strict Nash equilibria under most adaptive procedures is immediate). The authors provide a detailed explanation for the particular class of dynamics that they study. These dynamics are richer and more sophisticated than the original fictitious play model, in particular because they incorporate active (and costly) experimentation into the model. The results do little to support the use of equilibrium analysis, as the necessary conditions for even local stability are strong. This work has done little so far to inform users of game models, because it does not tell you what the players are doing when they are not at an equilibrium.[19]

The theoretical models of adaptive dynamics obtain some of their power by assuming that all agents follow the same learning process. When the learning rules are naive or decision makers are myopic, however, the question arises as to whether someone who is aware of how others are acting can exploit agents. Imposing a stability notion of this sort will not change the behavior of systems already in equilibrium. It could provide operational limits to the range of adaptive rules that could persist in a strategic environment.

Models in which learning is described by Bayes rule lead to surprisingly powerful results. Kalai and Lehrer [37] present one of the few analyses of repeated games in which players are not myopic. The agents in these models begin with beliefs about their opponents' strategies that are not necessarily correct, but are reasonable. Kalai and Lehrer [37] demonstrate that eventually beliefs of this kind, updated using observations of play and Bayes rule, will almost surely generate predictions of future play that are nearly correct. This conclusion depends on Bayesian updating, but not on any assumption about how players select their strategies. Suppose that an agent thinks that her opponent will play one strategy from a family. She can imagine the outcome that would arise following any conceivable strategy of her opponent. It is clear that if the family of possible strategies is finite, then after a finite length of time a history will be consistent with only one continuation. Kalai and Lehrer [37] demonstrate that this conclusion continues to be (approximately) true even when conjectures are not finitely supported. A corollary of the result is that if players optimize subject to their beliefs, then eventually play begins to look like equilibrium behavior. These results are limited in two respects. First, as in the social learning models, the asymptotic result may disguise the fact that agents' behavior is far from equilibrium for a long time. Second, Nachbar [48] has

[19] This work also identified the concept of self-confirming equilibrium (Fudenberg and Levine [26] and Kalai and Lehrer [38]). This notion weakens Nash equilibrium by taking into account the possibility that players may never be able to arrive at the common expectations needed to support a Nash equilibrium if they are unable to observe opponents' strategies in the course of a play of the game.

demonstrated, roughly, if one player is able to predict future play in non-trivial environments, then he will not have the ability to optimize.[20]

Several fundamental models of strategic interaction under incomplete information were developed in the seventies and early eighties. Substantive economic issues provided motivation for models of adverse selection, signaling, reputation, and limit pricing, to name a few, were motivated by. Game theoretical solution concepts developed in direct response to these models. The current class of learning models, proposed in response to deficiencies of equilibrium analysis, is often twice removed from economic problems. Save for some recent exceptions, they rarely come with economic applications. The research has, however, provided a new way to look at robustness of behavior in games, which leads to alternative theoretical explanations for choosing to concentrate on a particular outcome in games. It has also helped to organize experimental data.

5. CONCLUSIONS

One can summarize a significant subset of the results I have described as follows. If an agent starts with a *sensible* model, if the environment is *stationary*, if it is *costless* to obtain and process information, then *eventually* the agent learns enough about the environment to make optimal decisions. This summary suggests four broad questions one might wish to answer. What is a sensible starting point for a learning model? What happens if the environment is not stationary? How does integrating the costs associated with gathering and processing information influence learning? How do agents behave before they learn to make optimal decisions?

Models necessarily must specify what agents initially know and how they build on this knowledge.[21] If agents start with a suitable Bayesian prior

[20] Nachbar [48] shows that if an agent's strategy is known to be in a class small enough for the opponent to be able to predict future play, then the class will not be large enough to allow the agent to choose a best response to the opponent's strategy.

[21] I can do no better than follow Holland, Holyoak, Nisbett, and Thagard [35, page 4] and quote Peirce [50]:

> Suppose a being from some remote part of the universe, where the conditions of existence are inconceivably different from ours, to be presented with a United States Census Report — which is for us a mine of valuable inductions. ... He begins, perhaps, by comparing the ratio of indebtedness to deaths by consumption in countries whose names begin with different letters of the alphabet. It is safe to say that he would find the ratio everywhere the same, and thus his inquiry would lead to nothing .... The stranger to this planet might go on for some time asking inductive questions that the Census would faithfully answer without learning anything except that certain conditions were independent of others.... Nature is a far vaster and less clearly arranged repertoire of facts than a census report; and if men had not come to it with special aptitudes for guessing right, it may well be doubted whether in the ten or twenty thousand years that they may have existed their greatest mind would have attained the amount of knowledge which is actually possessed by the lowest idiot.

over possible models, which they refine by repeatedly observing information consistent with their priors, then strong results are available that suggest that these agents will ultimately be able to predict the future. Models that depend upon agents relying solely on their prior to process and refine new information are central to insightful strategic models of reputation formation and bargaining. Yet, judging from the results of simple experimental tests of Bayes Rule and the absence of applications of the learning literature, learning models that assume agents begin with Bayesian priors assume too much. Choosing an appropriate starting point for an analysis depends on many considerations. My next comments suggest considerations that should be taken into account in formulating models.

Agents learn by using past experiences to obtain a better understanding of their current and future problems. Circumstances (the relationship between actions and outcome) change. Past experience may not be directly relevant to current actions. While stationary environments provide a useful starting point for analysis, limiting analysis to stationary environments focuses attention on learning rules and theoretical questions that may not be relevant to non-stationary environments. In non-stationary environments learning is necessarily an ongoing process. Agents may never reach optimal decisions. Locking into a single action or following historically popular actions are extreme and generally inappropriate ways to behave.

Agents face a large number of simple, familiar interactions between friends and family. Cooperation is rational in these situations. If agents generalize their experience to other, less familiar and potentially less friendly environments, then cooperative behavior may arise in strategic interactions more broadly than folk-theorem arguments would suggest. Studies of learning by analogy in non-stationary environments might explain the prevalence of non-strategic behavior in strategic settings (and the appearance of agents who exploit trusting behavior). Yet it is not clear how to define optimal learning in non-stationary environments. It may be necessary to investigate other criteria, for example, long-run survival or some version of evolutionary fitness, to investigate the quality of various decision rules.[22]

Conlisk [15] makes a convincing argument that economic models pay little attention to how hard it is to make decisions. The standard (for economists) definition of rationality requires that agents automatically solve problems that are, in fact, beyond the abilities of any agent. While the

[22] Gilboa and Schmeidler [32] and Sarin and Vahid [55] (in a decision-theoretic environment) and Gale and Rosenthal [31] (in a strategic environment) study rules suited for non-stationary environments. Thus far, they have studied the behavior of these models only in stationary environments. These papers do not study procedures (like sequential hypothesis testing) that allow agents to test whether there has been a basic structural change in the environment that would cause an agent to abandon past experience entirely.

models that I have discussed depend on agents having limited information, most did not ask how the agents decide to acquire the information that they receive, nor how hard these agents must work to process this information. Allowing for decision costs can be done in simple, analytically tractable ways. Doing so might provide answers to interesting questions that go beyond current models. For example, intuition suggests that people are more likely to make good decisions when they have experience making similar decisions, when the stakes are high,[23] and when the decision is easy. The models usually identify what a good decision is. Analysis, however, almost always concentrates on whether agents with sufficient experience making the same decision converge to a good decision. Rarely do papers explicitly discuss the sensitivity of decisions to the stakes of the decision. The models do not consider what makes a decision problem easy. Nor has there been much energy identifying the tradeoffs between different ways of acquiring or using information. Once one admits that agents might make bad decisions because they are inexperienced, ignorant, or face difficult problems, the possibility of policy designed to lessen these inefficiencies (by simplifying information or training or by designing institutions that place less onerous information processing burdens on participants) arises.[24] It also provides a framework for understanding things that are purposefully designed to exploit learning deficiencies.[25]

A central question in the literature on learning is: When does learning lead, asymptotically, to good results? No one would expect the answer to be always or never. My reading of the literature suggests that the answer frequently is "not for a very long time." Hence the results that support the proposition that agents make optimal decisions asymptotically tell us little about the kinds of decisions people normally make. If (under the appropriate conditions) learning leads to good outcomes in a realistic amount of time, then we must refine our models in order to obtain the faster rates of convergence that we observe in experiments.[26] On the other hand, the models provide broad scope for the theoretical assertion that learning does not lead to optimal decisions in a realistic length of time. This observation suggests the need to identify properties of the entire path

[23] If high stakes are correlated with high stress, psychological factors may lower the quality of decision making.

[24] Clever people who design products knew enough to make the gasoline tanks of cars designed for unleaded gasoline too small for the nozzles of leaded gas. Hospital connections come in different sizes, lowering the chance that a patient will receive ether instead of oxygen.

[25] Many forms of advertisement (especially commercials targeted at children) and the arrangement of items on supermarket shelves are two common examples of information carefully targeted to take advantage of consumers' limited capacity to process information.

[26] Alternatively, we may attribute laboratory results to experimental designs that lessen learning problems faced by real-world agents.

of actions. As it undermines a traditional argument in favor of optimizing models, it provides support for further study of non-optimizing models as well.

Intelligently designed institutions perform well even if individual participants are poorly informed or boundedly rational. The literature on learning could identify institutions that lead to good outcomes either because learning is easier or faster in those settings or because outcomes are not sensitive to poor decisions that agents may make.[27] Although some experimental (Gode and Sunder [33]) and field (McAfee and McMillan [43]) research exists on how to design markets, learning models have the potential to add a new dimension to study of mechanism design.[28] Similarly, understanding the conditions under which optimal learning does not arise might provide a framework for understanding how clever actors could create environments explicitly designed to exploit learning deficiencies.

REFERENCES

1. P. Aghion, P. Bolton, C. Harris, and B. Jullien, Optimal Learning by Experimentation, *Rev. Econ. Stud.* **58** (1991), 621–654.
2. B. Arthur, On designing economic agents that behave like human agents, *J. Evol. Econ.* **3** (1993), 1–22.
3. A. Banerjee, A simple model of herd behavior, *Quart. J. Econ.* **107** (1992), 797–817.
4. A. Banerjee and D. Fudenberg, "Word-of-Mouth Learning," Harvard, 1995.
5. J. Bergin and B. Lipman, Evolution with state-dependent mutations, *Econometrica* **64** (1996), 943–956.
6. D. Berry and B. Fristedt, "Bandit Problems," Chapman and Hall, London, 1985.
7. S. Bikhchandani, D. Hirshleifer and I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades, *J. Polit. Econ.* **100** (1992), 992–1026.
8. K. Binmore and L. Samuelson, Evolutionary stability in repeated games played by finite automata, *J. Econ. Theory* **57** (1992), 278–305.
9. A. Blume, Y.-G. Kim, and J. Sobel, Evolutionary stability in games of communication, *Games Econ. Behavior* **5** (1993), 547–575.
10. T. Børgers, On the relevance of learning and evolution to economic theory, *Econ. J.* **106** (1996), 1374–1385.
11. G. Brown, Iterated solution of games by fictitious play, *in* "Activity Analysis of Production and Allocation" (T. C. Koopmans, Ed.), Wiley, New York, 1951.
12. A. Cabrales, Adaptive dynamics and the implementation problem with complete information, *J. Econ. Theory* **86** (1999), 159–184.
13. A. Cabrales and J. Sobel, On the limit points of discrete selection dynamics, *J. Econ. Theory* **57** (1992), 407–4419.

[27] These comments echo those of Marimon [62].

[28] Cabrales [12] identifies mechanisms in the implementation literature that do and do not select globally stable outcomes of adaptive dynamic adjustment processes.

14. P. Churchland, "The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain," MIT Press, Cambridge, 1995.
15. J. Conlisk, Why bounded rationality, *J. Econ. Lit.* **34** (1996), 669–700.
16. E. Dekel and S. Scotchmer, On the evolution of optimizing behavior, *J. Econ. Theory* **57** (1992), 392–406.
17. G. Ellison, Learning, local interaction, and coordination, *Econometrica* **61** (1993), 1047–1072.
18. G. Ellison and D. Fudenberg, Rules of thumb for social learning, *J. Political Econ.* **101** (1993), 612–643.
19. G. Ellison and D. Fudenberg, Word-of-mouth communication and social learning, *Quart. J. Econ.* **110** (1995), 93–125.
20. W. K. Estes, Individual behavior in uncertain situations, *in* "Decision Processes" (R. M. Thrall, C. H. Coombs, and R. L. Davis, Eds.), Wiley, New York, 1954.
21. D. Foster and P. Young, Stochastic evolutionary game dynamics, *Theoret. Popul. Biol.* **38** (1990), 219–232.
22. D. Foster and P. Young, On the nonconvergence of fictitious play in coordination games, *Games Econ. Behavior* **25** (1998), 79–96.
23. D. Fudenberg and C. Harris, Evolutionary dynamics with aggregate shocks, *J. Econ. Theory* **57** (1991), 420–441.
24. D. Fudenberg and D. Kreps, Learning and mixed equilibria, *Games Econ. Behavior* **5** (1993), 320–367.
25. D. Fudenberg and D. Kreps, Learning in extensive-form games, I: Self-confirming equilibria, *Games Econ. Behavior* **8** (1995), 20–55.
26. D. Fudenberg and D. Levine, Self-confirming equilibrium, *Econometrica* **61** (1993), 523–545.
27. D. Fudenberg and D. Levine, Consistency and cautious fictitious play, *J. Econ. Dynam. Control* **19** (1995), 1065–1089.
28. D. Fudenberg and D. Levine, Learning in games: where do we stand, *Europ. Econ. Rev.* **42** (1998), 631–639.
29. D. Fudenberg and E. Maskin, Evolution and cooperation in noisy repeated games, *Amer. Econ. Rev.* **80** (1990), 274–279.
30. D. Gale, What have we learned from social learning? *Europ. Econ. Rev.* **40** (1996), 617–628.
31. D. Gale and R. Rosenthal, Experimentation, imitation, and stochastic stability, *J. Econ. Theory* **84** (1999), 1–40.
32. I. Gilboa and D. Schmeidler, Case based optimization, *Games Econ. Behavior* **15** (1996), 1–26.
33. D. Gode and S. Sunder, Allocative efficiency of markets with zero-intelligence traders, *J. of Polit. Econ.* **101** (1993), 119–137.
34. F. Gul, Rationality and coherent theories of strategic behavior, *J. Econ. Theory* **70** (1996), 1–31.
35. J. Holland, K. Holyoak, R. Nisbett, and P. Thagard, "Induction," MIT Press, Cambridge, 1989.
36. E. Hutchins, "Cognition in the Wild," MIT Press, Cambridge, 1996.
37. E. Kalai and E. Lehrer, Rational learning leads to Nash equilibrium, *Econometrica* **61** (1993), 1019–1045.
38. E. Kalai, and E. Lehrer, Subjective equilibria in repeated games, *Econometrica* **61** (1993), 1231–1240.
39. M. Kandori, Evolutionary game theory in economics, *in* "Advances in Economics and Econometrics" (D. Kreps and K. Wallis, Eds.), Cambridge University Press, Cambridge, UK, 1997.

40. M. Kandori, G. Mailath, and R. Rob, Learning to play equilibria in games with stochastic perturbations, *Econometrica* **61** (1993), 29–56.
41. A. Kirman, Ants, rationality, and recruitment, *Quart. J. Econ.* **108** (1993), 137–156.
42. R. Marimon, Learning from learning in economics, *in* "Advances in Economics and Econometrics" (D. Kreps and K. Wallis, Eds.), Cambridge University Press, Cambridge, UK, 1997.
43. R. P. McAfee and J. McMillan, Analyzing the airwaves auction, *J. Econ. Persp.* **10** (1996), 159–176.
44. P. Milgrom and J. Roberts, Rationalizability, learning, and equilibrium in games with strategic complementarities, *Econometrica* **58** (1990), 1255–1277.
45. P. Milgrom and J. Roberts, Adaptive and sophisticated learning in normal form games, *Games Econ. Behavior* **3** (1991), 82–100.
46. K. Miyasawa, On the convergence of the learning process in a 2 non-zero sum two-person game, Princeton University, 1961.
47. D. Monderer and L. Shapley, Fictitious play property, *J. Econ. Theory* **68** (1996), 258–265.
48. J. Nachbar, Prediction, optimization, and rational learning in games, *Econometrica* **65** (1997), 275–309.
49. G. Nöldeke and L. Samuelson, An evolutionary analysis of backward and forward induction, *Games Econ. Behavior* **5** (1992), 425–454.
50. C. S. Peirce, A theory of probable inference, *in* "Collected Papers, II" (C. Hartshorne, P. Weiss, and A. Burks, Eds.) MIT Press, Cambridge, MA, 1958.
51. A. Robson, Efficiency in evolutionary games: Darwin, Nash, and the secret handshake, *J. Theoret. Biol.* **144** (1990), 379–396.
52. J. Robinson, An iterative method of solving a game, *Ann. Math.* **54** (1951), 296–301.
53. M. Rothschild, A two-armed bandit theory of market pricing, *J. Econ. Theory* **9** (1974), 185–202.
54. L. Samuelson and J. B. Zhang, Evolutionary stability in asymmetric games, *J. Econ. Theory* **57** (1992), 363–391.
55. R. Sarin and F. Vahid, Payoff assessment without probabilities: A simple dynamic model of choice, *Games Econ. Behavior* **28** (1999), 294–309.
56. L. Shapley, Some topics in two-person games, *in* "Advances in Game Theory, Annals of Mathematical Studies," (M. Dresher, L. Shapley, and A. Tucker, Eds.), Vol. 5, pp. 1–28, Princeton University Press, Princeton, 1964.
57. H. Simon, A comparison of game theory and learning theory, *Psychometrika* **21** (1956), 267–272.
58. D. Smallwood and J. Conlisk, Product quality in markets where consumers are imperfectly informed, *Quart. J. Econ.* **93** (1979), 1–23.
59. L. Smith and P. Sørensen, Informational herding and optimal experimentation, MIT, 1998.
60. L. Smith and P. Sørensen, Pathological outcomes of observational learning, *Econometrica* **68** (2000), 371–398.
61. X. Vives, How fast do rational agents learn? *Rev. Econ. Stud.* **60** (1993), 329–347.
62. K. Wärneryd, Cheap talk, coordination, and evolutionary stability, *Games Econ. Behavior* **5** (1993), 532–546.