

Arne Ryde Memorial Lectures 2002

*Strategic Learning and
Its Limits*

H. PEYTON YOUNG

Johns Hopkins University

and

University of Oxford

OXFORD
UNIVERSITY PRESS

Contents

<i>Acknowledgements</i>	xi
1. The Interactive Learning Problem	1
2. Reinforcement and Regret	10
2.1. Reinforcement learning	10
2.2. Learning in stationary environments	16
2.3. Criteria of performance	18
2.4. Regret	19
2.5. Regret matching	21
2.6. Realized payoffs	22
2.7. The logic of regret matching	25
3. Equilibrium	29
3.1. Forms of equilibrium	29
3.2. Examples	32
3.3. A generalization of correlated equilibrium	34
3.4. Learning coarse correlated equilibrium	36
3.5. Concepts of convergence	39
4. Conditional No-Regret Learning	43
4.1. Conditional versus unconditional regret	43
4.2. Blackwell's approachability theorem	45
4.3. Eliminating conditional regret	51
4.4. Simple rules minimizing conditional regret	54
4.5. A generalization of Blackwell's Theorem	57
4.6. Summary	60
5. Prediction, Postdiction, and Calibration	62
5.1. Prediction of an unknown process	62
5.2. An impossibility theorem of Oakes	64
5.3. Random forecasting rules	68

5.4.	Foster's forecasting rule	72
5.5.	Calibrated forecasting and correlated equilibrium	74
6.	Fictitious Play and Its Variants	76
6.1.	Predictive learning rules	76
6.2.	Smoothed fictitious play	80
6.3.	Better versus best reply	83
6.4.	Finite memory and inertia	84
6.5.	Convergence for weakly acyclic games	86
7.	Bayesian Learning	91
7.1.	The inference problem	91
7.2.	An example	92
7.3.	Strategies and beliefs	95
7.4.	Optimality and equilibrium	99
7.5.	Uncertainty and robustness	103
7.6.	An impossibility theorem	106
7.7.	Further implications	110
8.	Hypothesis Testing	113
8.1.	Cognitive learning theory	113
8.2.	Cognitive learning in games	114
8.3.	The structure of hypothesis testing	116
8.4.	Naive hypothesis testing	119
8.5.	Dynamics of hypothesis testing	122
8.6.	Learning Nash equilibrium	129
8.7.	Hypothesis testing: the general case	131
8.8.	Models, hunches, and beliefs	137
8.9.	Convergence in probability	139
8.10.	Learning to predict	142
9.	Conclusion	144
	References	149
	Index	159

Preface

This essay is based on a series of lectures given in the autumn of 2002 at the University of Lund. The event was sponsored by the Arne Ryde Foundation in memory of Arne Ryde, a former doctoral student at Lund. Previous lectures in the series, notably those by Thomas J. Sargent and by Kumara Velupillai, have been concerned with models of bounded rationality and their application to learning economic equilibrium. The present essay is in the same tradition, but the domain of application is to learning in games rather than to macroeconomic phenomena.

My purpose is not to provide a complete and up-to-date survey of the theoretical learning literature. Nor do I attempt to evaluate experimental evidence that seeks to test the empirical validity of different classes of rules. The recent pace of developments—both theoretical and experimental—would make this a futile exercise. Instead, I suggest a conceptual framework to help organize our thinking about strategic learning, and to highlight some of the theoretical achievements to date. This framework emphasizes the amount of *information* required to implement different types of learning rules, criteria for evaluating their *performance*, and alternative, sometimes novel, notions of *equilibrium* to which they converge. I also stress the limits of what can be achieved: for a given type of game and a given amount of information there may exist *no* learning procedure that satisfies certain reasonable criteria of performance and convergence. In sum, my goal is to provide a primer that delineates what we know, what we would like to know, and the limits of what we can know, when we try to learn about a system that is composed of other learners.

The Interactive Learning Problem

EQUILIBRIUM is as central to the study of social systems as it is to the analysis of physical phenomena. In the physical world equilibrium results from a balancing of forces. In societies it results from a balancing of intentions. In a physical system, particles are in equilibrium when they do not deviate from a given position or stable trajectory. In a social system, individuals' intentions are in equilibrium when no one wants to deviate from his intended behavior given the intentions of others.

Classical mechanics is based on a theory of forces and how they operate when a physical system is either in or out of equilibrium. In economics and the other social sciences there is no comparable framework for modeling out-of-equilibrium behavior by groups of individuals. This is not to say that economists are unaware of the need for such a theory; indeed the issue of how market prices and demands come into equilibrium is a long-standing problem in economics. But this is a different question from the one considered here, which is how groups of individuals, interacting strategically, adapt their behavior to the observed behavior of others.

Perhaps the nearest thing we have to a fully developed framework for analyzing this question is Bayesian decision theory. Suppose that individuals can imagine all possible future states of the world, both in and out of equilibrium. Suppose further that they can imagine all possible changes in behavior—by all the other individuals in society—over all possible sequences of states. As conditions unfold, they update their beliefs and adapt their behaviors to optimize expected future payoffs. If their beliefs put positive

probability on the strategies their opponents are actually using, then beliefs and behaviors will gradually come into alignment, and equilibrium, or something close to it, will eventually obtain.

Although this high rationality view is accepted in much of economic theory, it is not free of difficulties. First, it requires that everyone be able to anticipate correctly the behaviors that their opponents are in fact using. Secondly, it requires that individuals optimize at each point in time as events unfold. Both of these assumptions presume an enormous amount of sophistication and reasoning power on the part of the participants, because optimization in the present requires computing expected discounted payoffs over all conceivable futures. Even if we believe that individuals are capable of making such calculations in principle, the question is whether—even in principle—all possible futures can actually be anticipated.

The reason why one might worry about this is that social systems have a peculiar feature that is not generally found in physical systems. A social system consists of individuals who are learning about a process in which others are learning. The system is self-referential. Learning the true state of the system is therefore quite unlike learning the values of parameters that govern a physical process, for example, or even the parameters that describe a social process that is external to the observer. When the observer is a part of the system, the act of learning changes the thing to be learned. It is therefore unclear whether there exist learning rules of any degree of complexity that can solve this problem consistently except in special cases.¹ And it is certainly unclear whether the problem can be solved using simple rules that bear some resemblance to actual learning behavior in humans.

¹ Binmore (1987, 1990) was among the first to call attention to this problem explicitly.

The difficulties posed by interactive learning are common to many areas of economics, and economists have long been sensitive to their potential implications for the attainment of equilibrium. A central question in general equilibrium theory, for example, is whether there exist natural price adjustment dynamics that converge to a competitive equilibrium from out-of-equilibrium conditions. It has long been known that some natural mechanisms—e.g. price changes in the direction of aggregate excess demands—do not work (Scarf, 1960; Sonnenschein, 1972). More generally, Saari and Simon (1985) have demonstrated that any convergent price mechanism must depend on virtually the entire matrix of marginal excess demands of every commodity with respect to its own price as well as the price of every other commodity. This puts an enormous informational burden on the adjustment mechanism, and implies that truly decentralized, convergent mechanisms are hard to come by.

Similar issues arise in macroeconomics, where the question is whether agents can learn rational expectations equilibrium from out-of-equilibrium conditions. There is a large literature on this question that is surveyed from a variety of perspectives by Sargent (1993), Grandmont (1998), and Evans and Honkapohja (2001). In the simplest version of this framework, a real-valued variable—say next period's price level—is determined by current and historical price levels and also by people's *expectations* about the price level next period. The question is whether a fixed point of the expectations mapping can be discovered through an adaptive learning process. Sargent exhibits various low dimensional learning rules—including least-squares estimation—that have essentially this property.

Grandmont points out, however, that convergence in such a system can be very sensitive to the nature of the learning process and the degree to which current expectations influence future outcomes. On the one hand, if

agents are sufficiently uncertain about the dynamics governing the process and do not rule out divergent trends that show up in small past deviations, then their expectations that the system *might* be unstable generate instability with near certainty. If, on the other hand, agents rule out divergent trends and *expect* stability, then expectations converge and the system is in fact stable. In a word, the expectations formation process is self-fulfilling. This means, however, that the dynamic behavior of the system is not determined by objective conditions, but by endogenous expectations about its behavior. As Grandmont suggests, this can be interpreted as a kind of “uncertainty principle” for learning in macroeconomic environments.

Many of these issues are also central to game theory, where the interaction between expectations and behaviors is even more starkly evident. Here, as in the case of macroeconomic dynamics, there are two competing schools of thought. One school points optimistically to particular classes of games that can in fact be learned by simple updating procedures, such as learning minimax equilibria by fictitious play (see Chapter 6). This school also points to the fact that any game can be learned by Bayesian methods provided that the priors are sufficiently aligned to begin with (Kalai and Lehrer, 1993).

The other school views the situation more pessimistically. Its adherents point out that simple learning procedures tend to work only in special cases, such as zero-sum games, potential games, and so forth. Moreover, they are unimpressed by the fact that Bayesian methods lead to learning if the priors are sufficiently aligned to begin with. First, Bayesian reasoning makes extreme demands on the computational capacities of the players, and hence is not a plausible model of decision making by human beings. Secondly, even if one is willing to overlook the empirical implausibility of this approach, the amount of coordination required to learn equilibrium effectively assumes some

notion of pre-equilibrium in the prior beliefs. Hence it does not solve the learning problem in a robust sense. (These matters are treated in Chapter 7.)²

The purpose of these lectures is to survey results on both sides of the question. On balance, I shall give a little more weight to the positive side, though not in a way that will please all of the positivists. I begin by examining simple forms of learning behavior that depend only on realized payoffs and simple statistical summaries of historical data. These include forms of reinforcement and no regret learning, which turn out to be closely connected. Some of these rules converge in long-run average behavior to correlated forms of equilibrium, including standard correlated equilibrium and variants thereof (see Chapters 2–5). An important property of these learning rules is that they are *robust*, that is, they require no prior knowledge of the opponents' payoffs or strategies. Significantly, however, they do not yield Nash equilibrium behavior except under special circumstances.

In Chapter 6 we study fictitious play and its variants. These can be viewed as primitive forms of Bayesian learning in which players use simple statistical models based on past evidence and choose best (or almost best) replies given the evidence. Smoothed versions of fictitious play have long-run average behavior that is similar to that of no regret learning, but they do not converge to Nash equilibrium except in special cases.

We then examine the case in which players are very sophisticated in order to see how much can be gained from a high rationality perspective (Chapter 7). The major positive result here is due to Kalai and Lehrer (1993), who show that Bayesian learning leads to Nash equilibrium

² Our concern here is with the learnability of Nash equilibrium in games with finite action spaces. If one allows infinite action spaces, there exist games in which one cannot even determine in finite time whether a Nash equilibrium exists, let alone learn it by a decentralized process (Rabin, 1957; Prasad, 1991, 1997).

behavior provided the players' prior beliefs capture the set of actual play paths with positive probability (the "absolute continuity" condition).

One limitation of this result is that absolute continuity assumes away a significant amount of the uncertainty inherent in the situation. In effect, it presupposes that players already have partial knowledge of their opponents' strategies, which allows them to eliminate many possibilities *ex ante* (Nachbar, 1997, 2001, 2003). This assumption is obviously problematic in situations where players are completely ignorant of their opponents' payoffs, for then they have no handle on the strategies that their opponents might be using.³

To illustrate this point concretely, consider the following two-person game in which each player has two strategies (Left and Right) and the payoffs are in goods:

The Soda Game		
	L	R
L	Coke, Coke	Sprite, Seven-Up
R	Seven-Up, Sprite	Pepsi, Pepsi

Assume that the players have von Neumann Morgenstern utilities for these goods, and they know their own payoffs, but they know nothing about the opponent's payoffs, except insofar as they may be revealed through repeated play. (In fact they may not even know the distribution from which the payoffs are drawn.) We shall call this an *uncertain game*, in contrast with a game of incomplete information, where the distribution of payoffs is generally assumed to be common knowledge. (Borrowing from Knight's terminology, we might describe the latter as a *risky game*.)

³ There is a parallel literature on the limits of Bayesian learning in stochastic general equilibrium environments. See in particular Blume, Bray, and Easley, 1982; Marimon, 1997; Blume and Easley, 1998; Sandroni, 2000.

Uncertain games present severe challenges to strategic learning. To illustrate, suppose that we are the row player in the Soda Game, and we have observed the following pattern of play through the first eleven periods (our period-by-period payoffs are shown on the top line):

	0	0	0	1	1	0	0	0	1	0	0	
Row	L	R	L	L	R	R	L	R	R	R	R	?
Col	R	L	R	L	R	L	R	L	R	L	L	?

What theory or model of the situation should guide our own choice of action next period? Notice that without more information about the opponent we do not even know what *kind* of a game we are facing. If both of us prefer “dark” drinks to “light” drinks or the other way around, it is a coordination game with three equilibria (two pure and one mixed). But if one of us prefers dark and the other prefers light, it is a game like matching pennies with a unique mixed equilibrium.

Now suppose that we are given some information about the distribution of the opponents’ payoffs. To be specific, suppose that the entries in the payoff matrix are determined by independent draws from a normal distribution. (This determines the game once and for all; the payoffs are not redrawn in each period.) Assume that both players are rational and Bayesian, and that each has a prior over the opponent’s strategy space that is guided by the commonly known payoff distribution. It can be shown that, under any pair of priors, the players will fail to learn Nash equilibrium with positive probability (Foster and Young, 2001). It follows, in particular, that there are *no* priors that satisfy the absolute continuity condition.

This result shows that rational learning has its limits: even when players are perfectly rational and arbitrarily forward looking there may be no priors that permit them to learn Nash equilibrium behavior. Similar negative results have been pointed out by Grandmont (1998) in the

context of macroeconomic learning, but in some respects the situation for repeated games is worse. Whereas Grandmont argues that *some* ways of updating expectations lead to instability in the neighborhood of equilibrium, the impossibility result described above says that, in some repeated games, learning by rational agents may fail to discover equilibrium no matter how their expectations are updated.

Quite apart from these theoretical limits to the “high rationality” framework, I would argue that Bayesian learning is too demanding in a computational sense to be taken seriously as a model of human decision making. Indeed, Savage himself drew attention to its implausibility when people need to make choices in complex environments. The Bayesian approach, said Savage, is summarized in the adage “look before you leap.” However, another adage also needs to be borne in mind: “you can cross that bridge when you come to it.” Savage summarized the trade-off between these perspectives as follows: “One must indeed look before he leaps, in so far as the looking is not unreasonably time-consuming and otherwise expensive; but there are innumerable bridges one cannot afford to cross, unless [one] happens to come to them” (Savage, 1954, p. 16).

In Chapter 8 we propose an alternative learning framework that allows for some forward-looking behavior, but also leaves some bridges to be crossed at a later opportunity. In this approach, players adopt simple probabilistic models of their opponents’ behavior, which they hold temporarily until the model can be tested against data. (This is the look-before-you-leap aspect.) If a model is found to be reasonably consistent with the data it is retained; otherwise the player rejects it and adopts a new model that is perhaps influenced by experience but not fully determined by it. (This is the bridge-crossing aspect.)

This type of learning behavior is similar to statistical hypothesis testing. It also has many of the features of cog-

nitive learning models in psychology, which seek to explain how subjects learn in complex environments (e.g. Rumelhart, 1980; Arthur, 1994). It can be shown that, as long as players use powerful tests and their responses are close to being optimal, this type of learning leads to Nash equilibrium behavior in any game. Moreover, it works even when the players have no knowledge (let alone common knowledge) of the opponents' payoffs, hence it is robust in a statistical sense.

In summary, there exist relatively simple, payoff-based rules that lead to variants of correlated equilibrium. In general they converge only in a long-run average sense, not in period-by-period behaviors. Furthermore, except for particular classes of games, these rules do not converge (even on average) to Nash equilibrium behavior. By contrast, Bayesian learning with full rationality does lead to Nash equilibrium behavior, but only if the priors are suitably aligned to begin with. This form of learning is therefore not robust; it can also be very complex to implement. Hence in my view it does not solve the learning problem in a satisfactory way. However, by introducing a small degree of randomness in the formation of beliefs and in the responses to those beliefs—in other words by backing off from rationality just a bit—one obtains robust learning procedures that are both simple to implement and lead to Nash equilibrium in general finite games.