

8

Behavioural Game Theory: Thinking, Learning and Teaching*

Colin F. Camerer,¹ Teck-Hua Ho and Juin Kuan Chong

Introduction

Game theory is a mathematical system for analysing and predicting how humans behave in strategic situations. Standard equilibrium analyses assume that all players: (1) form beliefs based on an analysis of what others might do (strategic thinking); (2) choose the best response given those beliefs (optimization); and (3) adjust best responses and beliefs until they are mutually consistent (equilibrium).

It is widely accepted that not every player behaves rationally in complex situations, so assumptions (1) and (2) are sometimes violated. For explaining consumer choices and other decisions, rationality may still be an adequate approximation even if a modest percentage of players violate the theory. But game theory is different: the players' fates are intertwined. The presence of players who do not think strategically or optimize, even if there are very few such players, can change what rational players should do. As a result, what a population of players is likely to do when some are not thinking strategically and optimizing can only be predicted by an analysis that uses the tools of (1)–(3) but accounts for bounded rationality as well, preferably in a precise way.¹

An alternative way to define the equilibrium condition (3) is that players are never surprised when the strategies of other players are revealed. Defined

* This research was supported by NSF grants SBR 9730364, SBR 9730187 and SES-0078911. Thanks to many people for helpful comments on this research, particularly Caltech colleagues (especially Richard McKelvey, Tom Palfrey and Charles Plott), Mónica Capra, Vince Crawford, John Duffy, Drew Fudenberg, John Kagel, members of the MacArthur Preferences Network, our research assistants and collaborators Dan Clendenning, Graham Free, David Hsia, Ming Hsu, Hongjai Rhee and Xin Wang, and seminar audience members and referees (especially two for this book) too numerous to mention. Dan Levin gave the shooting-ahead military example of sophistication. Dave Cooper, Ido Erev and Guillaume Frechette wrote helpful emails.

this way, it seems unlikely that equilibrium (3) is reached instantaneously in one-shot games. Indeed, in the modern view, equilibrium should be thought of as the limiting outcome of an unspecified learning or evolutionary process that unfolds over time.² In this view, equilibrium is the *end* of the story of how strategic thinking, optimization, and equilibration (or learning) work, not the beginning (one-shot) or the middle (equilibration).

This chapter has three goals. First we develop an index of bounded rationality which measures players' steps of thinking and uses one parameter to specify how heterogeneous a population of players is. Coupled with best response, this index makes a unique statistical prediction of behaviour in any one-shot game. Second, we develop a learning algorithm (called a self-tuning EWA – experience-weighted attraction) to compute the path of equilibration. The algorithm generalizes both fictitious play and reinforcement models, and has shown greater empirical predictive power than those models in many games (adjusting for complexity, of course). Consequently, the self-tuning EWA can serve as an empirical device for finding the behavioural resting point as a function of the initial conditions. Third, we show how the index of bounded rationality and the learning algorithm can be extended to understand repeated game behaviours such as reputation building and strategic teaching. The idea is to present new, procedurally-rational models inspired by data, in the spirit of Werner Güth's (for example, Güth, 2000) many, many contributions of a similar kind.

Our approach is guided by three stylistic principles: precision, generality and empirical discipline. The first two are standard desiderata in game theory; the third is a cornerstone in experimental economics.

Precision

Because game theory predictions are sharp, it is not hard to spot likely deviations and counter-examples. Until recently, most of the experimental literature consisted of documenting deviations (or successes) and presenting a simple model, usually specialized to the game at hand. The hard part is to distil the deviations into an alternative theory that is as precise as standard theory and can be applied widely. We favour specifications that use one or two free parameters to express crucial elements of behavioural flexibility, because people are different. We also prefer to let data, rather than our intuition, specify parameter values.³

Generality

Much of the power of equilibrium analyses, and their widespread use, comes from the fact that the same principles can be applied to many different games, using the universal language of mathematics. Widespread use of the language creates a dialogue that sharpens theory and accumulates worldwide know-how. Behavioural models of games are also meant to be general, in the sense that the models can be applied to many games with minimal

customization. The insistence on generality is common in economics, but is not universal. Many researchers in psychology believe that behaviour is so context-specific that it is impossible to have a common theory that applies to all contexts. Our view is that we cannot know whether general theories fail until they are applied widely. Showing that customized models of different games fit well does not mean there isn't a general theory waiting to be discovered that is even better.

It is noteworthy that, in the search for generality, the models we describe below typically fit dozens of different data sets, rather than just one or two. The number of subject periods used when games are pooled is usually several thousand. This does not mean the results are conclusive or unshakeable. It just illustrates what we mean by a general model. The emphasis on explaining many data sets with a single model is also meant to create a higher standard for evaluating models of learning (and limited cognition as well). When an investigator reports only a single experiment and a learning model customized to explain those data, it is hard to take such models seriously until their generality has been explored by applying them to other games.

Empirical discipline

Our approach is heavily disciplined by data. Because game theory is about people (and groups of people) thinking about what other people and groups will do, it is unlikely that pure logic alone will tell us what will happen. As the physicist Murray Gell-Mann said, 'Think how hard physics would be if particles could think.' It is even harder if we do not watch what 'particles' do when interacting. Or, as Thomas Schelling (1960, p. 164) wrote, 'One cannot, without empirical evidence, deduce what understandings can be perceived in a nonzero-sum game of maneuver any more than one can prove, by purely formal deduction, that a particular joke is bound to be funny.'

Our insistence on empirical discipline is shared by others, past and present. Von Neumann and Morgenstern (1944) thought that 'the empirical background of economic science is definitely inadequate . . . it would have been absurd in physics to expect Kepler and Newton without Tycho Brahe, – and there is no reason to hope for an easier development in economics'. Fifty years later, Eric Van Damme (1999) had a similar thought:

Without having a broad set of facts on which to theorize, there is a certain danger of spending too much time on models that are mathematically elegant, yet have little connection to actual behavior. At present our empirical knowledge is inadequate and it is an interesting question why game theorists have not turned more frequently to psychologists for information about the learning and information processes used by humans.

The data we use to inform theory are experimental. Laboratory environments provide crucial control of what players know, when they move, and

what their payoffs are, which is crucial because game-theoretic predictions are sensitive to these variables (see Crawford, 1997). As in other lab sciences, the idea is to use lab control to sort out which theories work well and which do not, then use them later to help understand patterns in naturally-occurring data. In this respect, behavioural game theory resembles data-driven fields such as labor economics or finance more than does analytical game theory. The large body of experimental data accumulated since the 1980s (and particularly over the late 1990s early 2000s; see Camerer, 2003) is a treasure trove that can be used to sort out which simple parametric models fit well.

While the primary goal of behavioural game theory models is to make accurate predictions when equilibrium concepts do not, it can also circumvent two central problems in game theory: refinement and selection. Because we replace the strict best-response (optimization) assumption with stochastic better-response, all possible paths are part of a (statistical) equilibrium. As a result, there is no need to apply subgame perfection or propose belief refinements (to update beliefs after zero-probability events where Bayes' rule is useless). Furthermore, with plausible parameter values, the thinking and learning models often solve the long-standing problem of selecting one of several Nash equilibria, in a statistical sense, because the models make a unimodal statistical prediction rather than predicting multiple modes. Therefore, while the thinking-steps model generalizes the concept of equilibrium, it can also be *more* precise (in a statistical sense) when equilibrium is imprecise (see Lucas, 1986).⁴

We shall make three remarks before proceeding. First, while we do believe the thinking, learning and teaching models in this chapter do a good job of explaining some experimental regularity parsimoniously, many other models are actively being explored.⁵ The models in this chapter illustrate what most other models also strive to explain, and how they are evaluated.

The second remark is that these behavioural models are shaped by data from game experiments, but are intended for eventual use in areas of economics where game theory has been applied successfully. We shall return to a list of potential applications in the conclusion, but to whet the reader's appetite, here is a preview. Limited thinking models might be useful in explaining price bubbles, speculation and betting, competition neglect in business strategy, simplicity of incentive contracts, and persistence of nominal shocks in macroeconomics. Learning might be helpful for explaining evolution of pricing, institutions and industry structure. Teaching can be applied to repeated contracting, industrial organization, trust-building, and policy-makers setting inflation rates.

The third remark is about how to read this long chapter. The second and third sections, on learning and teaching, are based on published research and an unpublished paper introducing the one-parameter self-tuning EWA approach. We chose some examples to highlight what the models do and how they compare to other models, and also use this opportunity to comment briefly on

methodology. The first section, on the cognitive hierarchy thinking model, is newer. Our 2002 working paper has many additional results. A short paper (Camerer *et al.*, 2003) reports some brief results.

We put all three into one paper to show the ambitions and sweep of behavioural game theory. It is no longer the case that game theory is a broad enterprise that has solved all the hard problems, and behavioural game theory has just nibbled at some minor anomalies, with no underlying theoretical principles. Formal behavioural models, sharply honed on data, have been proposed in all the central issues in non-cooperative game theory – How will people play a new game? How do they learn? Do they behave differently when they are playing against others repeatedly? A natural question is how the models of thinking, learning and teaching fit together. The short answer is that the learning model is a mathematical special case of teaching in which all players are adaptive rather than sophisticated. The thinking and teaching models are not designed to fit together, since the cognitive hierarchy model is designed to apply to one-shot games (where planning for future rounds does not matter) and teaching is clearly for repeated games. One could presumably unify the models by mapping increasing steps of strategic thinking into increasing degrees of sophistication, but we have not done so. Or calling the thinking steps ‘rules’ and allowing players to learn in the domain of rules is a way of unifying the two (for example, Stahl, 1996). Since the models are so parsimonious there is no great saving in degrees of freedom by unifying them, but it would be important, both scientifically and practically, to know if there is a close link.

A cognitive hierarchy thinking model and bounded rationality measure

The cognitive hierarchy (CH) model is designed to predict behaviour in one-shot games and to provide initial conditions for models of learning. The model uses an iterative process which formalizes Selten’s (1998, p. 421) intuition that ‘the natural way of looking at game situations . . . is not based on circular concepts, but rather on a step-by-step reasoning procedure’.

We begin with notation. Strategies have numerical attractions that determine the probabilities of choosing different strategies through a logistic response function. For player i , there are m_i strategies (indexed by j) which have initial attractions denoted $A_i^j(0)$. Denote i ’s j th strategy by s_i^j , chosen strategies by i and other players (denoted $-i$) in period t as $s_i(t)$ and $s_{-i}(t)$, and player i ’s payoffs of choosing s_i^j by $\pi_i(s_i^j, s_{-i}(t))$.

A logit response rule is used to map attractions into probabilities:

$$P_i^j(t+1) = \frac{e^{\lambda A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda A_i^k(t)}} \quad (8.1)$$

where λ is the response sensitivity.⁶

We model thinking by characterizing the number of steps of iterated thinking that subjects do, and their decision rules.⁷ In the CH model some players, using zero steps of thinking, do not reason strategically at all. (Think of these players as being fatigued, clueless, overwhelmed, uncooperative, or simply more willing to make a random guess in the first period of a game and learn from subsequent experience than to think hard before learning.) We assume that zero-step players randomize equally over all strategies.

Players who do one step of thinking *do* reason strategically. What exactly do they do? We assume they are ‘over-confident’ – though they use one step, they believe others are all using zero steps. Proceeding inductively, players who use K steps think all others use zero to $K - 1$ steps.

It is useful to ask why the number of steps of thinking might be limited. One answer comes from psychology. Steps of thinking strain ‘working memory’, where items are stored while being processed. Loosely speaking, working memory is a hard constraint. For example, most people can remember only about five to nine digits when shown a long list of digits (though there are reliable individual differences, correlated with reasoning ability). The strategic question: ‘If she thinks he anticipates what she will do what should she do?’ is an example of a recursive ‘embedded sentence’ of the sort that is known to strain working memory and produce inference and recall mistakes.⁸

Reasoning about others might also be limited, because players are not certain about other players’ payoffs or degree of rationality. Why should players think others are rational? After all, adherence to optimization and instant equilibration is a matter of personal taste or skill. But whether other players do the same is a guess about the world (and iterating further, a guess about the contents of another player’s brain or a firm’s boardroom activity).

The key challenge in thinking steps models is pinning down the frequencies of K -step thinkers, $f(K)$. The constraint on working memory suggests that the relative fraction of $K - 1$ step thinkers doing one more step of thinking, $f(K)/f(K - 1)$, should be declining in K . For example, suppose the relative proportions of 2-step thinkers and 3-step thinkers is proportional to $1/3$, the relative proportions of 5- and 6-step thinkers is proportional to $1/6$, and in general $f(K)/f(K - 1) \propto 1/K$. This axiom turns out to imply that $f(K)$ has a Poisson distribution with mean and standard deviation τ (the frequency of level K types is $f(K) = \frac{e^{-\tau} \tau^K}{K!}$). Then τ is an index of the degree of bounded rationality in the population.

The Poisson distribution has three appealing properties: it has only one free parameter (τ); since Poisson is discrete it generates ‘spikes’ in predicted distributions reflecting individual heterogeneity (other approaches do not⁹); and for sensible τ values the frequency of step types is similar to the frequencies estimated in earlier studies (see Stahl and Wilson, 1995; Ho *et al.*, 1998; and Nagel *et al.*, 1999). When we compare Poisson-constrained distributions $f(K)$ to a 7-parameter free distribution, with $f(0), f(1), \dots, f(7)$ each a

separate parameter, the Poisson distribution fits only about 1 per cent worse in log-likelihood terms in most data sets (see Camerer *et al.*, 2002b), so it is a very close approximation to a freer distribution.

Figure 8.1 shows four Poisson distributions with different τ values. Note that there are substantial frequencies of steps 0–3 for τ around one or two. There are also very few higher-step types, which is plausible if the limit on working memory has an upper bound.

Modelling heterogeneity is important, because it allows the possibility that not every player is rational. The few studies that have looked carefully found fairly reliable individual differences, because a subject's step level or decision rule is fairly stable across games (Stahl and Wilson, 1995; Costa-Gomes *et al.*, 2001). Including heterogeneity can also improve learning models by starting them off with enough persistent variation across people to match the variation seen among actual people.

To make the CH model precise, assume players know the absolute frequencies of players at lower levels from the Poisson distribution. But since they do not imagine higher-step types, there is missing probability. They must adjust their beliefs by allocating the missing probability in order to compute sensible expected payoffs to guide choices. We assume players divide the correct relative proportions of lower-step types by $\sum_{c=1}^{K-1} f(c)$ so the adjusted frequencies maintain the same relative proportions but add up to one.

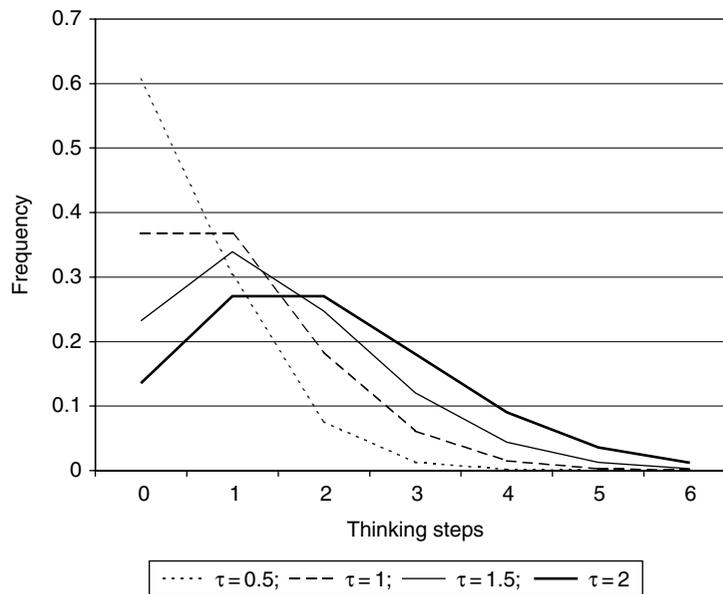


Figure 8.1 Poisson distributions for various τ

Given this assumption, players using $K > 0$ steps are assumed to compute expected payoffs given their adjusted beliefs, and use those attractions to determine choice probabilities according to

$$A_i^j(0|K) = \sum_{h=1}^{m-i} \pi_i(s_i^j, s_{-i}^h) \times \left\{ \sum_{c=0}^{K-1} \left[\frac{f(c)}{\sum_{c=0}^{K-1} f(c)} \times P_{-i}^h(1|c) \right] \right\} \quad (8.2)$$

where $A_i^j(0|K)$ and $P_i^j(1|c)$ are the attraction of level K in Period 0, and the predicted choice probability of lower level c in Period 1.

As a benchmark, we also fit quantal response equilibrium (QRE), defined by:

$$A_i^j(0|K) = \sum_{h=1}^{m-i} \pi_i(s_i^j, s_{-i}^h) \times P_{-i}^h(1) \quad (8.3)$$

$$P_i^j(1) = \frac{e^{\lambda A_i^j(0)}}{\sum_{h=1}^{m_i} e^{\lambda A_i^h(0)}} \quad (8.4)$$

When λ goes to infinity, QRE converges to the Nash equilibrium. QRE is closely related to a thinking-steps model in which K -step types are 'self-aware' and believe there are other K -step types, and τ goes to infinity.

Fitting the model

As a first pass, the thinking-steps model was fitted to data from three studies in which players made decisions in matrix games once for each without feedback (a total of 2,558 subject-games).¹⁰ Within each of the three data sets, a common λ was used, and best-fitting τ values were estimated, both separately for each game and fixed across games (maximizing log likelihood). Our working paper (Camerer *et al.*, 2002b) contains many more games and more model comparisons.

Table 8.1 reports τ values for each game separately, common τ and λ from the thinking-steps model, and measures of fit for the thinking model and QRE – the log likelihood (LL) (which can be used to compare models) and the mean of the squared deviations (MSD) between predicted and actual frequencies.

QRE fits a little worse than the thinking model in all three data sets.¹¹ This is a hint that a self-awareness model is not necessarily more accurate than one with extreme over-confidence.

Estimated values of τ are quite variable in the Stahl and Wilson data, but fairly consistent in the others.¹² In the latter two sets of data, estimates are clustered around one and two, respectively. Imposing a common τ across games only reduces fit very slightly (even in the Stahl and Wilson games¹³). The fact that the cross-game estimates are the most consistent in the Costa-Gomes *et al.* (2001) games, which have the most structural variation among them, is also encouraging.

Table 8.1 Estimates of thinking model τ and fit statistics, three matrix game experiments

	<i>Stahl and Wilson (1995)</i>	<i>Cooper and Van Huyck (2003)</i>	<i>Costa-Gomes et al. (2001)</i>
<i>Game-specific τ estimates</i>			
Game 1	18.34	1.14	2.17
Game 2	2.26	1.04	2.21
Game 3	1.99	0.00	2.22
Game 4	4.56	1.25	1.44
Game 5	5.53	0.53	1.81
Game 6	1.70	0.80	1.58
Game 7	5.55	1.17	1.08
Game 8	2.03	1.75	1.94
Game 9	1.79		1.88
Game 10	8.79		2.66
Game 11	7.33		1.34
Game 12	21.46		2.30
Game 13			2.36
Common τ	8.44	0.81	2.22
Common λ	9.06	190.58	15.76
<i>Fit statistics (thinking steps model)</i>			
MSD (pooled)	0.0257	0.0135	0.0063
LL (pooled)	-1115	-1739	-555
<i>Fit statistics (QRE)</i>			
MSD (QRE)	0.0327	0.0269	0.0079
LL (QRE)	-1176	-1838	-599

Note: In Costa-Gomes *et al.* (2001) the games are labelled and 2b-2 \times 2, 3a-2 \times 2, 3b-2 \times 2, 4b-3 \times 2, 4c-3 \times 2, 5b-3 \times 2, 8b-3 \times 2, 9a-4 \times 2, 4a-2 \times 3, 4d-2 \times 3, 6b-2 \times 3, 7b-2 \times 3 and 9b-2 \times 4.

Furthermore, while the values of λ we estimate are often quite large, the overall frequencies the model predicts are close to the data. That means that a near-best-response model with a mixture of thinking steps can fit a little better than a QRE model, which assumes stochastic response but has only one 'type'. The heterogeneity may therefore enable modellers to use best-response calculations and still make probabilistic predictions, which is enormously helpful analytically. The large estimated values of λ also mean that λ can be set to a large value, or simply assume best-response of $K > 1$ -step types, which is a big computational saving and fits almost as well (Camerer *et al.* 2002a).

Figures 8.2 and 8.3 show how accurately the CH and Nash models fit the data from the three matrix-game data sets. In each figure, the data points are separate strategies from each of the games. Figure 8.2 shows that the data and

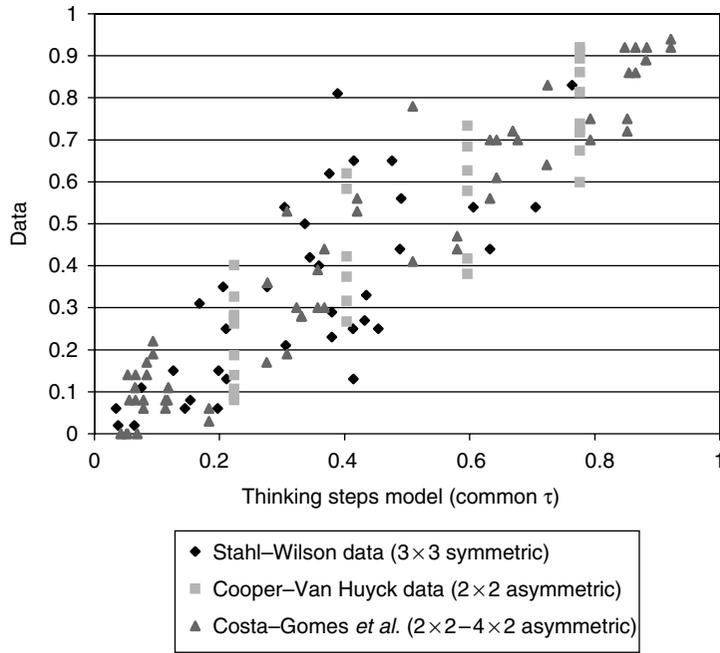


Figure 8.2 Fit of thinking-steps model to three games ($R^2 = 0.84$)

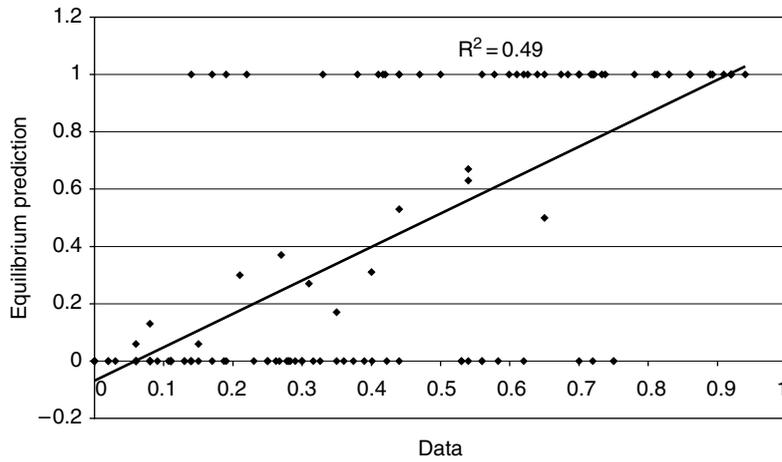


Figure 8.3 Nash equilibrium predictions versus data in three games

fits are reasonably good, while Figure 8.3 shows that the Nash predictions (which are often zero or 1, pure equilibria, are reasonably accurate, though not as close as the thinking-model predictions). Since τ is consistently around 1–2, the CH model with a single τ could be an adequate approximation to first-period behaviour in many different games. To see how far the model can take us, we investigated it in two other classes of games – games with mixed equilibria, and binary entry games. The next section describes results from entry games (see the Appendix on page 163 for details on mixed games).

Market entry games

Consider binary entry games in which there is capacity c (expressed as a fraction of the number of entrants). Each of many entrants decides simultaneously whether to enter or not. If an entrant thinks that fewer than c per cent will enter, s/he will enter; if s/he thinks more than c per cent will enter, s/he stays out.

There are three regularities in many experiments based on entry games such as this one (see Ochs, 1999; Seale and Rapoport, 1999; Camerer, 2003, ch. 7): (i) entry rates across different capacities c are closely correlated with entry rates predicted by (asymmetric) pure equilibria or symmetric mixed equilibria (that is, about c per cent of the people enter when capacity is c); (ii) players slightly over-enter at low capacities and under-enter at high capacities; and (iii) many players use noisy cut-off rules in which they stay out for most capacities below some cut-off c^* and enter for most higher capacities.

Let us apply the CH thinking model with best response. Step zero players enter half the time. This means that when $c < 0.5$, 1-step thinkers stay out, and when $c > 0.5$ they enter. Players doing 2 steps of thinking believe the fraction of zero steppers is $f(0)/(f(0) + f(1)) = 1/(1 + \tau)$. Therefore, they enter only if $c > 0.5$ and $c > \frac{0.5 + \tau}{1 + \tau}$, or when $c < 0.5$ and $c > \frac{0.5}{1 + \tau}$. To make this more concrete, suppose $\tau = 2$. Then 2-step thinkers enter when $c > 5/6$ and $1/6 < c < 0.5$. What happens is that more steps of thinking ‘iron out’ steps in the function relating c to overall entry. In the example, 1-step players are afraid to enter when $c < 1/2$. But when c is not too low (between $1/6$ and 0.5) the 2-step thinkers perceive room for entry because they believe the relative proportion of zero-steppers is $1/3$ and those players enter half the time. Two-step thinkers stay out for capacities between 0.5 and $5/6$, but they enter for $c > 5/6$ because they know half of the $(1/3)$ zero-step types will randomly stay out, leaving room even though 1-step thinkers always enter. Higher steps of thinking smooth out steps in the entry function even further.

The surprising experimental fact is that players can co-ordinate entry reasonably well, even in the first period. (‘To a psychologist’, Kahneman (1988) wrote, ‘this looks like magic.’) The thinking-steps model provides a possible explanation for this magic and can account for the other two regularities for

reasonable τ values. Figure 8.4 plots entry rates from the first block of two studies for a game similar to the one above (Sundali *et al.*, 1995; Seale and Rapoport, 1999). Note that the number of actual entries rises almost monotonically with c , and entry is above capacity at low c and below capacity at high c .

Figure 8.4 also shows the CH entry function $N(all|\tau)(c)$ for $\tau = 1.5$ and 2. Both functions reproduce monotonicity and the over- and under-capacity effects. The thinking-steps models also produces approximate cut-off rule behaviour for all higher thinking steps except two. When $\tau = 1.5$, step 0 types randomize, step 1 types enter for all c above 0.5, step 3–4 types use cut-off rules with one 'exception', and levels 5 and above use strict cut-off rules. This mixture of random, cut-off and near-cut-off rules is roughly what is observed in the data when individual patterns of entry across c are measured (see, for example, Seale and Rapoport, 1999).

The model can also be used to do some simple theory. For example, in Camerer *et al.*, 2002b (and Camerer, Ho and Chong, 2003) we show that the entry function is monotonic (that is, $N(all|\tau)(c)$ is increasing in c) if $1 + 2\tau < e^\tau$, or $\tau < 1.25$. (The same condition guarantees that the conditional

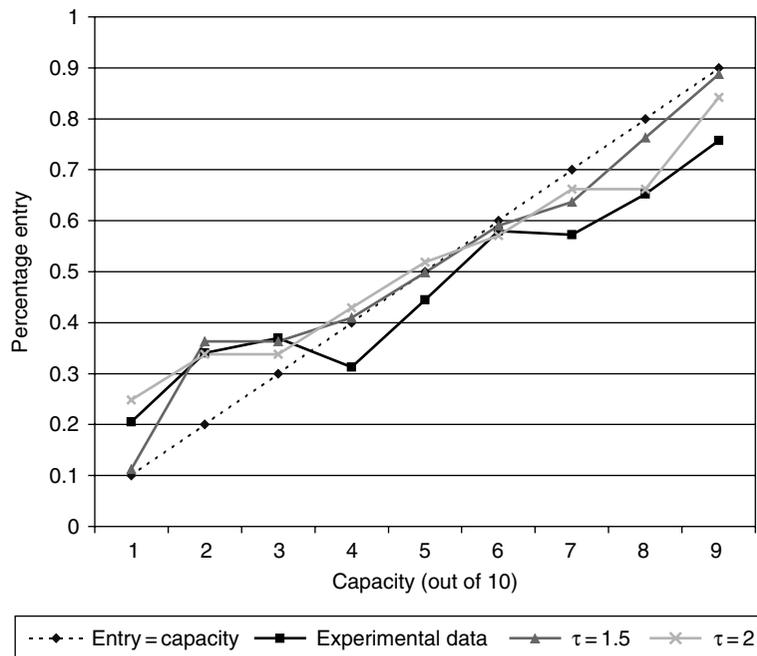


Figure 8.4 How entry varies with capacity (c), data and thinking-steps model

entry functions including only up to K -step players get increasingly close to the equilibrium entry as K rises.)

Thinking steps and cognitive measures

Since the CH model is a cognitive model, it gives an account of some treatment effects and shows how cognitive measures, such as response times and information acquisition, can be correlated with choices.

Belief-prompting

Some studies show that asking players for explicit beliefs about what others will do moves their choices, moving them closer to equilibrium (compared to a control in which beliefs are not prompted – see, for example, Costa-Gomes and Weizsacker, 2003). A simple example reported in Warglien *et al.* (1998) is shown in Table 8.2. Best-responding 1-step players think others are randomizing, so they will choose X , which pays 60, rather than Y , which has an expected payoff of 45. Higher-step players choose Y .

Without belief-prompting, 70 per cent of the row players choose X . When subjects are prompted to articulate a belief about what the column players will do, 70 per cent choose the dominance-solvable equilibrium choice Y . Croson (2000) reports similar effects. In experiments on beauty contest games, we found that prompting beliefs also reduced dominance-violating choices modestly. Schotter *et al.* (1994) found a related display effect – showing a game in an extensive-form tree led to more subgame perfect choices.

Belief-prompting can be interpreted as increasing all players' thinking by one step. To illustrate, assume that since step zeros are forced to articulate *some* belief, they move to step 1. Now they believe others are random so they choose X . Players previously using one or more steps now use two or more. They believe column players choose L so they choose Y . The fraction of X play is therefore because former zero-step thinkers now do one step of thinking. This is just one simple example, but the numbers match up reasonably well¹⁴ and it illustrates how belief-prompting effects could be accommodated within the thinking-steps model.

Similarly, Cooper and Kagel (2003b) report that two-person teams play signalling games more strategically (and transfer learning better) than individuals (though see Kocher and Sutter, fourth coming). This might be understood

Table 8.2 How belief-prompting promotes dominance-solvable choices by row players

Row move	Column player		Without belief prompting	With belief prompting
	L	R		
X	60,20	60,10	0.70	0.30
Y	80,20	10,10	0.30	0.70

Source: (Warglien *et al.*, 1998).

formally in terms of a model in which the highest-step player ‘teaches’ the lower-step player.

Information look-ups

Camerer *et al.* (1993), Costa-Gomes *et al.* (2001), Johnson *et al.* (2002), and Salmon (2003) measure directly the information subjects acquire in a game by putting payoff information in boxes which must be clicked open using a computer mouse. The order in which boxes are opened, and for how long they are open, gives a ‘subject’s-eye view’ of what players are looking at, and should be correlated with thinking steps. Indeed, Johnson *et al.* show that how much time players spend looking ahead to future ‘pie sizes’ in alternating-offer bargaining is correlated with the offers they make. Costa-Gomes *et al.* show that look-up patterns are correlated with choices that result from various (unobserved) decision rules in normal-form games. These correlations means that a researcher who simply knew what a player had looked at could, to some extent, forecast that player’s offer or choice. Both studies also showed that information look-up statistics helped to answer questions that choices alone could not.¹⁵

Summary

A simple cognitive hierarchy model of thinking steps attempts to predict choices in one-shot games and provide initial conditions for learning models. We propose a model which incorporate discrete steps of thinking, and the frequencies of players using different numbers of steps is Poisson-distributed with mean τ . We assume that players at level $K > 0$ cannot imagine players at their level or higher, but they understand the relative proportions of lower-step players and normalize them to compute expected payoffs. Estimates from three experiments on matrix games show reasonable fits for τ around 1–2, and τ is fairly regular across games in two of three data sets. Values of $\tau = 1.5$ also fits data from fifteen games with mixed equilibria and reproduces key regularities from binary entry games. The thinking-steps model also creates natural heterogeneity across subjects. When best response is assumed, the model generally creates ‘purification’ in which most players at any step level use a pure strategy, but a mixture results because of the mixture of players using different numbers of steps.

Learning

By the mid-1990s, it was well-established that simple models of learning could explain some movements in choice over time in specific game and choice contexts.¹⁶ Therefore, the issue is not whether simple models of learning can capture some aspects of movement in experimental data – that issue was well-settled (the answer is Yes) by the late 1990s. The bigger challenge taken up since then is to see how well a specific parametric model can

account for finer details of the equilibration process in a very wide range of games.

This section describes a one-parameter theory of learning in decisions and games called functional EWA (or self-tuning EWA for short; also called 'functional EWA' or 'EWA lite' to emphasize its simple functions). Self-tuning EWA predicts the time path of individual behaviour in any normal-form game. Initial conditions can be imposed or estimated in various ways. We use initial conditions from the thinking-steps model described in the previous section. The goal is to predict both initial conditions and equilibration in new games in which behaviour has never been observed, with minimal free parameters (the model uses one parameter, λ).

Parametric EWA learning: interpretation, uses and limits

Self-tuning EWA is a relative of the parametric model of learning called experience-weighted attraction (EWA) (Camerer and Ho, 1998, 1999). As in most theories, learning in EWA is characterized by changes in (unobserved) attractions based on experience. Attractions determine the probabilities of choosing different strategies through a logistic response function. For player i , there are m_i strategies (indexed by j) which have initial attractions denoted $A_i^j(0)$. The thinking steps model is used to generate initial attractions given parameter values τ and λ .

To avoid complications with negative payoffs, we rescale payoffs by subtracting by the minimum payoff so that rescale payoffs are always weakly positive. Define an indicator function $I(x, y)$ to be zero if $x \neq y$ and one if $x = y$. The EWA attraction updating equation is:

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{N(t-1)\phi(1-\rho) + 1} \quad (8.5)$$

and the experience weight (the 'EW' part) is updated according to $N(t) = N(t-1)\phi(1-\rho) + 1$.¹⁷ Notice that the term $[\delta + (1-\delta)I(s_i^j, s_i(t))]$ implies that a weight of one is put on the payoff term when the strategy being reinforced is the one the player chose ($s_i^j = s_i(t)$), but the weight on forgone payoffs from unchosen strategies ($s_i^j \neq s_i(t)$) is δ . (When forgone payoffs are not known exactly, averaging possible values or using historical rules can be used as proxies.¹⁸) Attractions are mapped into choice probabilities using a logit response function $P_i^j(t+1) = \frac{e^{\lambda A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda A_i^k(t)}}$ (where λ is the response sensitivity). The subscript i , superscript j , and argument $t+1$ in $P_i^j(t+1)$ are reminders that the model aims to explain every choice by every subject in every period.¹⁹

In implementing the model, we shall typically take strategies to be stage-game strategies. However, it is often likely that a strategy could be history-dependent or have some other form, which should be considered in future

work. Furthermore, transfer of learning across games is an interesting topic we have not explored (but see Cooper and Kagel, 2003a).

Each EWA parameter has a natural interpretation.

The parameter δ is the weight placed on forgone payoffs. Presumably it is affected by imagination (in psychological terms, the strength of counterfactual reasoning or regret, or in economic terms, the weight placed on opportunity costs and benefits), or reliability of information about forgone payoffs (Heller and Sarin, 2000).

The parameter ϕ decays previous attractions because of forgetting or, more interestingly, because agents are aware that the learning environment is changing and deliberately 'retire' old information (much as firms junk old equipment more quickly when technology changes rapidly).

The parameter ρ controls the rate at which attractions grow. When $\rho = 0$, attractions are weighted averages and grow slowly; but when $\rho = 1$ attractions cumulate. We originally included this variable because some learning rules used cumulation and others used averaging. It is also a rough way to capture the distinction in machine learning between 'exploring' an environment (low ρ), and 'exploiting' what is known by locking in to a good strategy (high ρ) (see, for example, Sutton and Barto, 1998).

The initial experience weight $N(0)$ is like a strength of prior beliefs in models of Bayesian belief learning. (Imposing $N(0) < \frac{1}{1+\phi(1-\rho)}$ guarantees that $N(t)$ is increasing, which is sensible.) It plays a minimal empirical role, so it is set to $N(0) = 1$ in our current work.

EWA is a hybrid of two widely-studied models, reinforcement and belief learning. In reinforcement learning, only payoffs from chosen strategies are used to update attractions and guide learning. In belief learning, players do not learn about which strategies work best; they learn about what others are likely to do, then use those updated beliefs to change their attractions and hence which strategies they choose (see Brown, 1951; Fudenberg and Levine, 1998). EWA shows that reinforcement and belief learning, which were often treated as being fundamentally different, are in fact related in a non-obvious way, because both are special kinds of reinforcement rules.²⁰ When $\delta = 0$, the EWA rule is a simple reinforcement rule.²¹ When $\delta = 1$ and $\rho = 0$, the EWA rule is equivalent to belief learning using weighted fictitious play.²²

It is important to be very clear about what the EWA formulation means and does. First, one thing EWA suggests is that general learning can be thought of as a splice of two different cognitive processes: in one process, strategies that are chosen are automatically (and fully) reinforced; and in the other, players think about the forgone payoffs they would have gained from choosing other strategies and reinforce them less strongly (with weight δ). Many kinds of behaviour are now attributable to the behaviour of two kinds of system (see, for example, Kahneman, 2003): one system is very fast, automatic, pattern-orientated and sometimes subconscious (similar to

our perceptual system), and can work in parallel (for example, people can perceive sounds and images at the same time); the other is slow, deliberate, conscious, invokes abstract rules and cognition (logic, arithmetic), and works serially. Put very roughly, δ represents the strength of the second, more deliberative, process which generates counterfactual answers to the question ‘Could I have done better with another strategy?’ Furthermore, seen in the dual-process light, reinforcement learning can be interpreted as suggesting that the second process does not start (which may be appropriate for animals with minimal cortical apparatus for deliberation, or for people under time pressure), while belief learning suggests that the second process completely overrides the rapid instinctive response of the first process.

Second, EWA provides a way of checking whether a simple model is *too simple*. Obviously, every model is a simplification that permits counterexamples. So making imperfect predictions is no reason to abandon a model. But imperfections *are* a reason to ask whether adding features to a model can improve its predictions. EWA is one way of doing precisely this.

Forgone payoffs are the fuel that runs EWA learning. They also provide an indirect link to ‘direction learning’ and imitation. In direction learning, players move in the direction of observed best response (Selten and Stöcker, 1986). Suppose players follow EWA but do not know forgone payoffs, and believe those payoffs are monotonically increasing between their choice $s_i(t)$ and the best response. If they also reinforce strategies near their choice $s_i(t)$ more strongly than strategies that are further away, their behaviour will look like direction learning. Imitating a player who is similar and successful can also be seen as a way of heuristically inferring high forgone payoffs from an observed choice and moving in the direction of those higher payoffs. However, this is probably not the whole story about ‘observational learning’, which is a fruitful area of research (see, for example, Duffy and Feltovich, 1999; Armantier, 2004).²³

The relationships of various learning rules can be shown visually in a cube showing configurations of parameter values (see Figure 8.5). Each point in the cube is a triple of EWA parameter values that specify a precise updating equation. The corner of the cube with $\phi = \rho = 0, \delta = 1$ is Cournot best-response dynamics. The corner $\rho = 0, \phi = \delta = 1$, is standard fictitious play. The vertex connecting these corners, $\delta = 1, \rho = 0$, is the class of weighted fictitious play rules (see, for example, Fudenberg and Levine, 1998). The vertices with $\delta = 0$ and $\rho = 0$ or 1 are averaging and cumulative choice reinforcement rules.

The biologist Francis Crick (1988) said, ‘in nature a hybrid is often sterile, but in science the opposite is usually true’. As Crick suggests, the point of EWA is not simply to show a surprising relationship among other models, but also to improve their fertility for explaining patterns in data by combining the best modelling ‘genes’. In reinforcement theories, received payoffs get the most weight (in fact, *all* the weight²⁴). Belief theories assume implicitly that

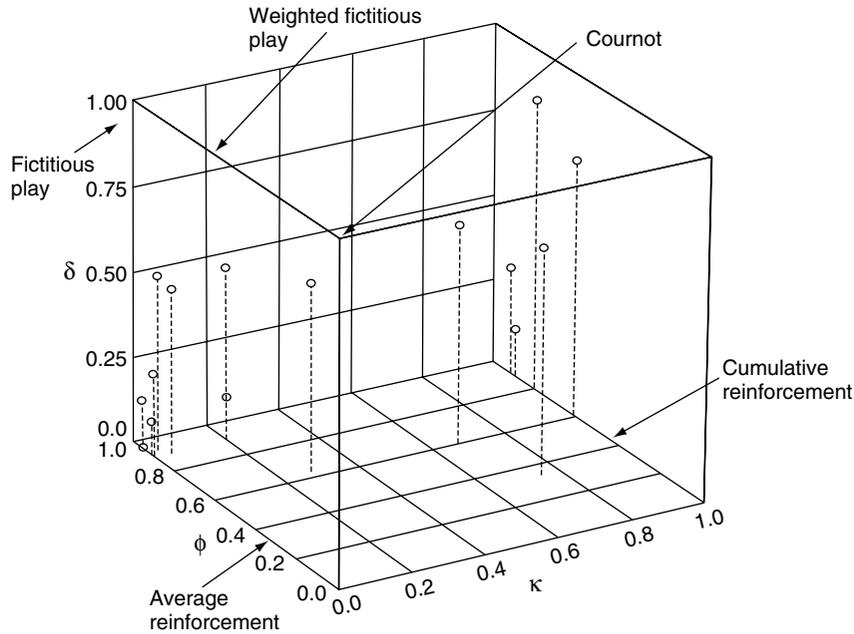


Figure 8.5 The EWA learning cube

forgone and received payoffs are weighted equally. Rather than assuming that one of these intuitions about payoff weights is right and the other is wrong, EWA allows *both* intuitions to be true. When $0 < \delta < 1$, received payoffs can gain more weight, but forgone payoffs also get some weight.

The EWA model has been estimated by ourselves and many others on about 40 data sets (see Camerer *et al.*, 2002). The hybrid EWA model predicts more accurately in many cases than do the special cases of reinforcement and weighted fictitious play, except in games with mixed-strategy equilibrium, where reinforcement does equally well.²⁵ It is extremely important to emphasize that, in our model estimation and validation, we *always* penalize the EWA model in ways that are known to generally make the *adjusted* fit *worse* if a model is too complex (that is, if the data are in fact generated by a simpler model).²⁶ Furthermore, econometric studies show that, if the data were generated by simpler belief or reinforcement models, then EWA estimates would generally correctly identify that fact for many games and reasonable sample sizes (see Cabrales and Garcia-Fontes, 2000; Salmon, 2001), although Wilcox (2003) finds that heterogeneity in all model parameters lowers the estimate of δ in mixed-equilibrium games (which may explain why low values of δ often fit well in these games). Since EWA is capable of identifying behaviour consistent with special cases, when it does not, then the hybrid parameter values are improving in fit.

Figure 8.5 also shows estimated parameter triples from 20 data sets. Each point is an estimate from a different game. If one of the special case theories is a good approximation to how people generally behave across games, estimated parameters should cluster in the corner or vertex corresponding to that theory. In fact, parameters tend to be sprinkled around the cube, although many (typically mixed-equilibrium games) cluster in the averaged reinforcement corner with low δ and ρ . The dispersion of estimates in the cube raises an important question: is there regularity in which games generate which parameter estimates? A positive answer to this question is crucial for predicting behaviour in brand new games.

This concern is addressed by a version of EWA, self-tuning EWA, which replaces free parameters with deterministic functions $\phi_i(t), \delta_i(t), \rho_i(t)$ of player i 's experience up to period t . These functions determine parameter values for each player and period. The parameter values are then used in the EWA updating equation to determine attractions, which then determine choices probabilistically. Since the functions also vary across subjects and over time, they have the potential to inject heterogeneity and time-varying 'rule learning', and to explain learning better than models with fixed parameter values across people and time. And since self-tuning EWA has only one parameter that must be estimated (λ),²⁷ it is especially helpful when learning models are used as building blocks for more complex models that incorporate sophistication (some players think others learn) and teaching, as we discuss in the section below.

The decay rate ϕ is sometimes interpreted as forgetting, an interpretation carried over from reinforcement models of animal learning. Certainly forgetting does occur, but the more important variation in $\phi_i(t)$ across games is probably a player's perception of how quickly the learning environment is changing. The function $\phi_i(t)$ should therefore 'detect change'. As in physical change detectors (for example, security systems or smoke alarms), the challenge is to detect change when it is really occurring, but not falsely mistake noise for change too often.

The core of the $\phi_i(t)$ change-detector function is a 'surprise index', which is the difference between other players' recent strategies and their strategies in previous periods. To make exposition easier, we describe the function for games with pure-strategy equilibria (suitably modified for games with mixed equilibria, as noted below). First define a history vector, across the other players' strategies k , which records the historical frequencies (including the last period) of the choices by other players of s_{-i}^k . The vector element $h_i^k(t)$ is $\frac{\sum_{\tau=1}^t I(s_{-i}^k, s_{-i}(\tau))}{t}$.²⁸ The recent 'history' $r_i^k(t)$ is a vector of 0s and 1s which has a 1 for strategy $s_{-i}^k = s_{-i}(t)$ and 0s for all other strategies s_{-i}^k (that is, $r_i^k(t) = I(s_{-i}^k, s_{-i}(t))$). The surprise index $S_i(t)$ simply sums up the squared deviations between the cumulative history vector $h_i^k(t)$ and the immediate recent history vector $r_i^k(t)$; that is, $S_i(t) \equiv \sum_{k=1}^{m-i} (h_i^k(t) - r_i^k(t))^2$. Note that this surprise index varies from zero (when the last strategy the other player chose

is the one they have always chosen before) to two (when the other player chose a particular strategy ‘for ever’ then suddenly switches to something brand new). The change-detecting decay rate is $\phi_i(t) = 1 - 0.5 \cdot S_i(t)$. Because $S_i(t)$ is between zero and two, ϕ is always (weakly) between one and zero.

The numerical boundary cases illuminate intuition: if the other player chooses the strategy s /he has always chosen before, then $S_i(t) = 0$ (player i is not surprised) and $\phi_i(t) = 1$ (player i does not decay the lagged attraction at all, since what other players did throughout is informative). If the other player chooses a new strategy that was never chosen before in a very long run of history, $S_i(t) = 2$ and $\phi_i(t) = 0$ (player i decays the lagged attraction completely and ‘starts over’). Note that, since the observed behaviour in period t is included in the history $h_i^k(t)$, $\phi_i(t)$ will typically not dip to zero. For example, if a player chose the same strategy for each of nine periods and a new strategy in period 10, then $S_i(t) = (0.9 - 0)^2 + (1 - 0.1)^2 = 2 \cdot 0.81$ and $\phi_i(t) = 1 - 0.5(2 \cdot 0.81) = 0.19$.

In games with mixed equilibria (and no pure equilibria), a player should expect other players’ strategies to vary. Therefore, if the game has a mixed equilibrium with W strategies that are played with positive probability (that is, W is the cardinality of the smallest support of any Nash strategy), the surprise index defines recent history over a window of the last W periods (for example, in a game with four strategies that are played in equilibrium, $W = 4$). Then $r_i^k(t) = \sum_{k=1}^{m-i} \left[\frac{\sum_{\tau=t-W+1}^t I_i^{(s^k, s-i(\tau))}}{W} \right]$,²⁹

A sensible property of $S_i(t)$ is that the surprisingness of a new choice should depend not only on how often the new choice has been chosen before, but also on how how variable previous choices have been. Incorporating this property requires ϕ to be larger when there is more dispersion in previous choices, which is guaranteed by squaring the deviations between current and previous history. (Summing absolute deviations between $r_i(t)$ and $h_i(t)$, for example, would not have this property.) If previously observed relative frequencies of strategy k are denoted f_k , and the recent strategy is h , then the surprise index is $(1 - f_h)^2 + \sum_{k \neq h} (f_k - 0)^2$. Holding f_h constant, this index is minimized when all frequencies f_k with $k \neq h$ are equal. In the equal- f_k case, the surprise index is $S_i(t) = (m_{-i} - 1)/m_{-i}$ and $\phi_i(t) = (m_{-i} + 1)/2m_{-i}$, which has a lower bound of 0.5 in games with large m_{-i} (many strategies).

The opposite case is when an opponent has previously chosen a single strategy in every period, and suddenly switches to a new strategy. In this case, $\phi_i(t)$ is $\frac{2t-1}{t^2}$. This expression declines gracefully towards zero as the string of identical choices up to period t grows longer. (For $t = 2, 3, 5$ and 10 , the $\phi_i(t)$ values are 0.75, 0.56, 0.36, and 0.19.) The fact that the ϕ values decline with t expresses the principle that a new choice is a bigger surprise (and should have an associated lower ϕ) if it follows a *longer* string of *identical* choices that are different from the surprising new choice. It also embodies the idea that dipping $\phi_i(t)$ too low is a mistake which should be avoided because it erases the history embodied in the lagged attraction. So $\phi_i(t)$ only

dips low when opponents have been choosing one strategy reliably for a very long time, and then switch to a new one.

Another interesting special case is when unique strategies have been played in every period up to $t - 1$, and another unique strategy is played in period t . (This is often true in games with large strategy spaces.) Then $\phi_i(t) = 0.5 + \frac{1}{2t}$, which starts at 0.75 and asymptotes at 0.5 as t increases.

The calculations above show that in the first few periods of a game, $\phi_i(t)$ will not dip much below 1 (because the t th period experience is included in the recent history $r_i(t)$ vector as well as being part of the cumulative history $h_i(t)$). But in these periods players often learn rapidly. Since it makes sense to start with a low value of $\phi_i(0)$ to express players' responsiveness in the first few periods, in the empirical implementation, we smooth the $\phi_i(t)$ function by starting at $\phi_i(0) = 0.5$, and gently blending in the updated values according to $\hat{\phi}_i(t) \equiv 0.5/t + (t - 1)\phi_i(t)/t$.

The other self-tuning EWA functions are less empirically important and interesting so we mention them only briefly. The function $\delta_i(t) = \phi_i(t)/W$. Dividing by W pushes $\delta_i(t)$ towards zero in games with mixed equilibria, which matches estimates in many games (see Camerer *et al.*, 2003).³⁰ Tying $\delta_i(t)$ to the change detector $\phi_i(t)$ means that chosen strategies are reinforced relatively strongly (compared to unchosen ones) when change is rapid. This reflects a 'status quo bias' or 'freezing' response to danger (which is virtually universal across species, including humans). Since $\rho_i(t)$ controls how sharply subjects lock in to choosing a small number of strategies, we use a 'Gini coefficient' – a standard measure of dispersion often used to measure income inequality – over choice frequencies.^{31,32}

Self-tuning EWA has three advantages. First, it is easy to use because it has only one free parameter (λ). Second, parameters in self-tuning EWA naturally vary across time and people (as well as across games), which can capture heterogeneity and mimic 'rule learning' in which parameters vary over time (see, for example, Stahl, 1996, 2000; Salmon, 1999). For example, if ϕ rises across periods from 0 to 1 as other players stabilize, players are effectively switching from Cournot-type dynamics to fictitious play. If δ rises from 0 to 1, players are effectively switching from reinforcement to belief learning. Third, it should be easier to theorize about the limiting behaviour of self-tuning EWA than about some parametric models. A key feature of self-tuning EWA is that, as a player's opponent's behaviour stabilizes, $\phi_i(t)$ goes toward 1 and (in games with pure equilibria) $\delta_i(t)$ does too. If $\rho = 0$, self-tuning EWA then automatically turns into fictitious play; and a lot is known about theoretical properties of fictitious play.

Self-tuning EWA predictions

In this section we compare in-sample fit and out-of-sample predictive accuracy of different learning models when parameters are estimated freely, and check whether self-tuning EWA functions can produce game-specific parameters similar to estimated values. We use seven games: games with unique

mixed strategy equilibrium (Mookerjee and Sopher, 1997); R&D (research and development) patent race games (Rapoport and Amaldoss, 2000); a median-action order statistic co-ordination game with several players (Van Huyck *et al.*, 1990); a continental-divide co-ordination game, in which convergence behaviour is extremely sensitive to initial conditions (Van Huyck *et al.*, 1997); a 'pots game' with entry into two markets of different sizes (Amaldoss and Ho, in preparation); dominance-solvable p-beauty contests (Ho *et al.*, 1998); and a price-matching game (called 'traveller's dilemma' by Capra *et al.*, 1999).

Estimation method

The estimation procedure for self-tuning EWA is sketched briefly here (see Ho *et al.*, 2001 for details). Consider a game where N subjects play T rounds. For a given player i of CH step-level c , the likelihood function of observing a choice history of $\{s_i(1), s_i(2), \dots, s_i(T-1), s_i(T)\}$ is given by:

$$\prod_{t=1}^T P_i^{s_i(t)}(t|c) \quad (8.6)$$

The joint likelihood function L of observing all players' choices is given by:

$$L(\lambda) = \prod_i^N \left\{ \sum_{c=1}^K f(c) \times \prod_{t=1}^T P_i^{s_i(t)}(t) \right\} \quad (8.7)$$

where K is set to a multiple of τ rounded to an integer. Most models are 'burnt in' by using first-period data to determine initial attractions. We also compare all models with burnt-in attractions with a model in which the thinking steps model from the previous section is used to create initial conditions and combined with self-tuning EWA. Note that the latter hybrid uses only two parameters (τ and λ) and does not use first-period data at all.

Given the initial attractions and initial parameter values,³³ attractions are updated using the EWA formula. Self-tuning EWA parameters are then updated according to the functions above and used in the EWA updating equation. Maximum likelihood estimation is used to find the best-fitting value of λ (and other parameters, for the other models) using data from the first 70 per cent of the subjects. Then the value of λ is frozen and used to forecast behaviour of the entire path of the remaining 30 per cent of the subjects. Payoffs were all converted to dollars (which is important for cross-game forecasting).

In addition to self-tuning EWA (one parameter), we estimated the parametric EWA model (five parameters), a belief-based model (weighted fictitious play, two parameters) and the two -parameter reinforcement models with payoff variability (Erev *et al.*, 1999; Roth *et al.*, 2000), and QRE.

Model fit and predictive accuracy in all games

The first question we ask is how well models fit and predict on a game-by-game basis (that is, parameters are estimated separately for each game). Some authors fix a set of parameters for several games and see how well predictions do based on those parameters. But it is impossible to know from such an exercise whether there is a *better* set of parameters, or how much the best-fitting parameters truly vary across games. By allowing parameter to vary across games we can both find out how well a single parameter would do (by looking at variation across games – if variation is low a single parameter would be fine, or we can restrict the parameter to have a common value and see how badly the fit degrades), and have some guidance as to which parameters fit best in which games.

An alternative method is to simulate the entire path of play and compare some statistics of the simulated path with statistics of the data (e.g., Roth and Erev 1995). We have done this in several publications and do not draw different conclusions from those drawn from the estimation method below (see, for example, Camerer *et al.*, 1998; Camerer, Hsia and Ho, 2002; and Camerer, 2003, ch. 6). Furthermore, note that if the statistic that is used is the *conditional* relative frequency (that is, frequencies conditioned on all actual histories), it makes no difference whether the conditioned histories are determined by the actual data, or paths are first simulated, then the paths which match the data histories are selected for conditional-frequency comparison. The simulation method will give exactly the same results as the method we use. Furthermore, if the frequencies that are compared are *not* conditioned on histories, then the model can be doing poorly on capturing some kind of history-dependence but appear to fit well on the unconditioned frequencies.³⁴

For out-of-sample validation we report both hit rates (the fraction of most-likely choices that are picked) and log likelihood (*LL*). (Keep in mind that these results forecast a holdout sample of subjects after model parameters have been estimated on an earlier sample and then 'frozen'. If a complex model is fitting better within a sample purely because of spurious overfitting, it will in fact predict less well out of sample.) Results are summarized in Table 8.3.

Across games, parametric EWA is as good as all other theories or better, judged by hit rate, and has the best *LL* in four games. Self-tuning EWA also does well on hit rate in six out of seven games. Reinforcement is competitive on hit rate in five games, and best in *LL* in two. Belief models are often inferior on hit rate and never best in *LL*. QRE clearly fits the worst.

Combining self-tuning EWA with a thinking steps model to predict initial conditions (rather than using the first-period data), a two-parameter combination is only a little worse in hit rate than self-tuning EWA, and slightly worse in *LL*.

Table 8.3 Out-of-sample accuracy of learning models

Game	Thinking +stEWA		stEWA		EWA		Weighted fict. play		Reinf. with PV		QRE	
	%Hit	LL	%Hit	LL	%Hit	LL	%Hit	LL	%Hit	LL	%Hit	LL
Cont'l divide (7)	45	-483	47	-470	47	- 460	25	-565	45	-557	5	-806
Med. action (14)	71	-112	74	-104	79	- 83	82	-95	74	-105	49	-285
p-BC (1)	8	-2119	8	-2119	6	- 2042	7	-2051	6	-2504	4	-2497
Pricing (0.8)	43	-507	46	-445	43	- 443	36	-465	41	-561	27	-720
Mixed games (21)	36	-1391	36	- 1382	36	-1387	34	-1405	33	-1392	35	-1400
Patents (18)	64	-1936	65	-1897	65	-1878	53	-2279	65	- 1864	40	-2914
Pot games (50)	70	-438	70	-436	70	-437	66	-471	70	- 429	51	-509
Pooled	50	-6986	51	- 6852	49	-7100	40	-7935	46	-9128	36	-9037
KS p-BC			6	-309	3	-279	3	-279	4	-344	1	-346

Notes: Sample sizes are 315, 160, 580, 160, 960, 1760, 739, 4674 (pooled), 80; The best fits for each game and criterion printed in bold; hit rates statistically indistinguishable from the best (by the McNemar test) also in bold. Numbers in parentheses after each game name in col. 1 are hit rates from a random model; comparing hit rates in the '%Hit' columns indicates how much better the models are doing than random prediction.

Source: Ho *et al.*, 2001.

The penultimate row of Table 8.3, 'pooled', shows results when a single set of common parameters is estimated for all games (except for game-specific λ ³⁵). If self-tuning EWA is capturing parameter differences across games effectively, it should predict especially accurately, compared to other models, when games are pooled. It does so: when all games are pooled, self-tuning EWA predicts out-of-sample better than other theories, by both statistical criteria.

Some readers of our functional EWA paper were concerned that, by searching across different specifications we may have overfitted the sample of seven games we reported. To check whether we did, we announced at conferences in 2001 that we would analyse all the data people sent us by the end of that year and report the results in a revised paper. Three samples were sent and we have analysed one so far – experiments by Kocher and Sutter (forthcoming) on p-beauty contest games played by individuals and groups. The KS results are reported in the last row of Table 8.3 ('KS p-BC'). The game is the same as the beauty contests we studied (except for the interesting complication of group decision-making, which speeds equilibration), so it is not surprising that the results replicate the earlier findings: belief and parametric EWA fit best by *LL*, followed by self-tuning EWA, and reinforcement and QRE models fit worst. This is a small piece of evidence that the solid performance of self-tuning EWA (while being worse than belief learning on these games) is not entirely caused by overfitting on our original seven-game sample.

Now we shall show predicted and relative frequencies for three games that highlight differences among models. In other games the differences are minor or hard to see with the naked eye.³⁶

Dominance-solvable games: beauty contests

In beauty contest games, each of n players chooses $x_i \in [0, 100]$. The average of their choices is computed and whichever player is closest to $p < 1$ times the average wins a fixed prize (see Nagel, 1999, for a review). The unique Nash equilibrium is zero. (The games get their name from a passage in Keynes about how the stock market is like a special beauty contest in which people judge who others will think is beautiful.) These games are a useful way to measure the steps of iterated thinking players seem to use (since higher steps will lead to lower number choices). Experiments have been run with exotic subject pools such as Ph.D.s and CEOs (Camerer, 1997), and in newspaper contests with very large samples (Nagel *et al.*, 1999). The results are generally robust, although specially-educated subjects (for example, professional game theorists) choose, not surprisingly, closer to equilibrium.

We analyse experiments run by Ho *et al.* (1998).³⁷ The data and relative frequencies predicted by each learning model are shown in Figure 8.6. Figure 8.6(a) shows that while subjects start around the middle of the distribution, they converge steadily downwards towards zero. By Period 5, half the subjects chose numbers 1–10.

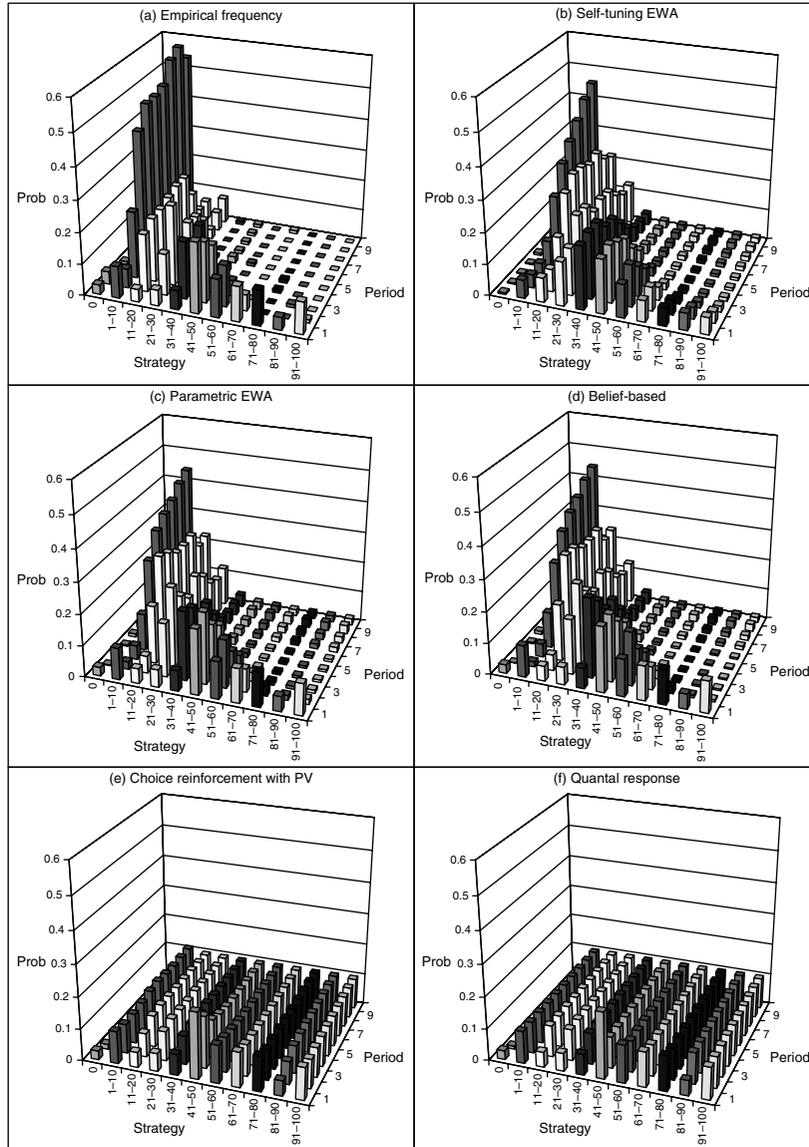


Figure 8.6 Predicted frequencies for p -beauty contest

The EWA, belief and thinking – self-tuning EWA model all capture the basic regularities although they underestimate the speed of convergence. (In the next section we add sophistication – some subjects know that others are learning and ‘shoot ahead’ of the learners by choosing lower

numbers – which improves the fit substantially.) The QRE model is a dud in this game, and reinforcement also learns far too slowly because most players receive no reinforcement.³⁸

Games with multiple equilibria: continental divide game

Van Huyck *et al.* (1997) studied a co-ordination game with multiple equilibria and extreme sensitivity to initial conditions, which we call the continental divide game (CDG). The payoffs in the game are shown in Table 8.4. Subjects play in cohorts of seven people. Subjects choose an integer from 1 to 14, and their payoff depends on their own choice and on the median choice of all seven players.

The payoff matrix is constructed so that there are two pure equilibria (at 3 and 12) which are Pareto-ranked (12 is the better one). Best responses to different medians are in bold. The best-response correspondence bifurcates in the middle: if the median starts at 7 virtually any sort of learning dynamics will lead players toward the equilibrium at 3. If the median starts at 8 or above, however, learning will eventually converge to an equilibrium of 12. Both equilibrium payoffs are shown in bold italics. The payoff at 3 is about half as much as at 12, so which equilibrium is selected has a large economic impact.

Figure 8.7 shows empirical frequencies (pooling all subjects) and model predictions.³⁹ The key features of the data are: bifurcation over time from choices in the middle of the range (5–10) to the extremes, near the equilibria at 3 and 12; and late-period choices are more clustered around 12 than around 3. There is also an extreme sensitivity to initial conditions (which is disguised by the aggregation across sessions in Figure 8.7(a)): namely, five groups had initial medians below 7 and all five converged toward the inefficient low equilibrium. The other five groups had initial medians above 7 and all five converged towards the efficient high equilibrium. This path-dependence shows the importance of a good theory of initial conditions (such as the thinking steps model). Because a couple of steps of thinking generates a distribution concentrated in the middle strategies 5–9, the thinking-steps model predicts that initial medians will sometimes be above the separatrix 7 and sometimes below. The model does not predict precisely which equilibrium will emerge, but it predicts that both high and low equilibria will sometimes emerge.

Notice also that strategies 1–4 are never chosen in early periods, but are chosen frequently in later periods. Strategies 7–9 are chosen frequently in early periods but rarely chosen in later periods. Like a sportscar, a good model should be able to capture these effects by ‘accelerating’ low choices quickly (going from zero to frequent choices in a few periods) and ‘braking’ midrange choices quickly (going from frequent choices to zero).

QRE fits poorly because it predicts no movement (it is not a theory of learning, of course, but simply a static benchmark that is tougher to beat

Table 8.4 Payoffs in 'continental divide' experiment

Choice	Median choice													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	45	49	52	55	56	55	46	-59	-88	-105	-117	-127	-135	-142
2	48	53	58	62	65	66	61	-27	-52	-67	-77	-86	-92	-98
3	48	54	60	66	70	74	72	1	-20	-32	-41	-48	-53	-58
4	43	51	58	65	71	77	80	26	8	-2	-9	-14	-19	-22
5	35	44	52	60	69	77	83	46	32	25	19	15	12	10
6	23	33	42	52	62	72	82	62	53	47	43	41	39	38
7	7	18	28	40	51	64	78	75	69	66	64	63	62	62
8	-13	-1	11	23	37	51	69	83	81	80	80	80	81	82
9	-37	-24	-11	3	18	35	57	88	89	91	92	94	96	98
10	-65	-51	-37	-21	-4	15	40	89	94	98	101	104	107	110
11	-97	-82	-66	-49	-31	-9	20	85	94	100	105	110	114	119
12	-133	-117	-100	-82	-61	-37	-5	78	91	99	106	112	118	123
13	-173	-156	-137	-118	-96	-69	-33	67	83	94	103	110	117	123
14	-217	-198	-179	-158	-134	-105	-65	52	72	85	95	104	112	120

Note: Best replies in bold.

Source: Van Huyck *et al.* (1997).

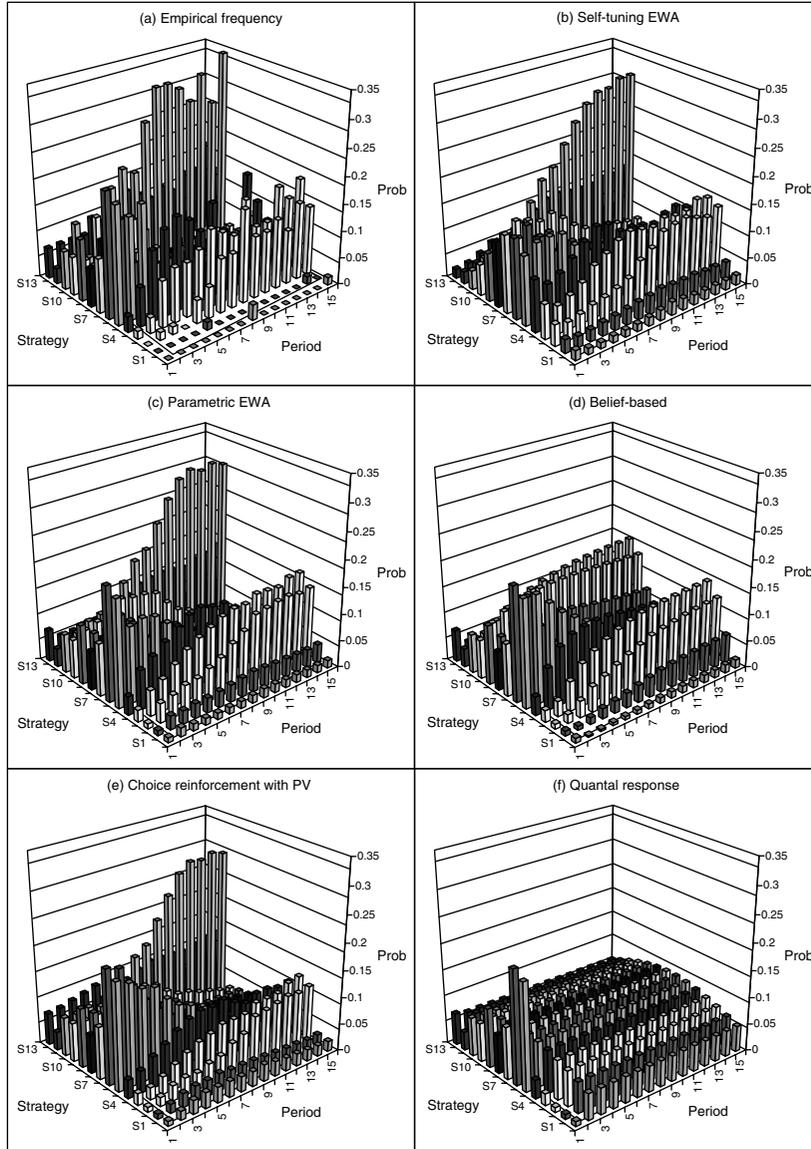


Figure 8.7 Predicted frequencies for continental divide

than Nash). Reinforcement with PV fits well. Belief learning does not reproduce the asymmetry between sharp convergence to the high equilibrium and flatter frequencies around the low equilibrium. The reason why is diagnostic of a subtle weakness in belief learning. Note from Table 8.4 that the

payoff gradients around the equilibria at 3 and 12 are exactly the same – choosing one number too high or low ‘costs’ \$0.02; choosing two numbers too high or low costs \$0.08, and so on. Since belief learning computes expected payoffs, and the logit rule means only differences in expected payoffs influence choice probability, the fact that the payoff gradients are the same means the spread of probability around the two equilibria must be the same. Self-tuning EWA, parametric EWA and the reinforcement models generate the asymmetry with low δ .⁴⁰

**Games with dominance-solvable equilibrium:
price-matching with loyalty**

Capra *et al.* (1999) studied a dominance-solvable price-matching game. In their game, two players simultaneously choose a price between 80 and 200. Both players earn the low price. In addition, the player who names the lower price receives a bonus of R and the player who names the higher price pays a penalty R . (If the prices they choose are the same, the bonus and penalty cancel and players just earn the price they named.) You can think of R as a reduced-form expression of the benefits of customer loyalty and word-of-mouth which accrue to the lower-priced player, and the penalty is the cost of customer disloyalty and switching away from the high-price firm. We like this game because price-matching is a central feature of economic life. These experiments can also, in principle, be tied to field observations in future work.

Their experiment used six groups of 9–12 subjects. The reward/penalty R had six values (5, 10, 20, 25, 50, 80). Subjects were rematched randomly.⁴¹

Figure 8.8 shows empirical frequencies and model fits for $R = 50$ (where the models differ most). A wide range of prices are named in the first round. Prices gradually fall, being 91–100 in Rounds 3–5, 81–90 in Rounds 5–6, and towards the equilibrium of 80 in later rounds.

QRE predicts a spike at the Nash equilibrium of 80.⁴² The belief-based model predicts the direction of convergence, but overpredicts numbers in the interval 81–90 and underpredicts choices of precisely 80. The problem is that the incentive in the traveller’s dilemma is to undercut the other player’s price by as little as possible. Players only choose 80 frequently in the last couple of periods; before those periods it pays to choose higher numbers.

EWA models explain the sharp convergence in late periods by cumulating payoffs and estimating $\delta = 0.63$ (for self-tuning EWA). Players who chose 80 while others named a higher price could have earned more by undercutting the other price, but weighting that higher forgone payoff by δ means their choice of 80 is reinforced more strongly, which matches the data.

Reinforcement with payoff variability has a good hit rate because the highest spikes in the graph often correspond with spikes in the data. But the graph shows that predicted learning is much more sluggish than in the data (that is, the spikes are not high enough). Because $\phi = 1$ and players are not

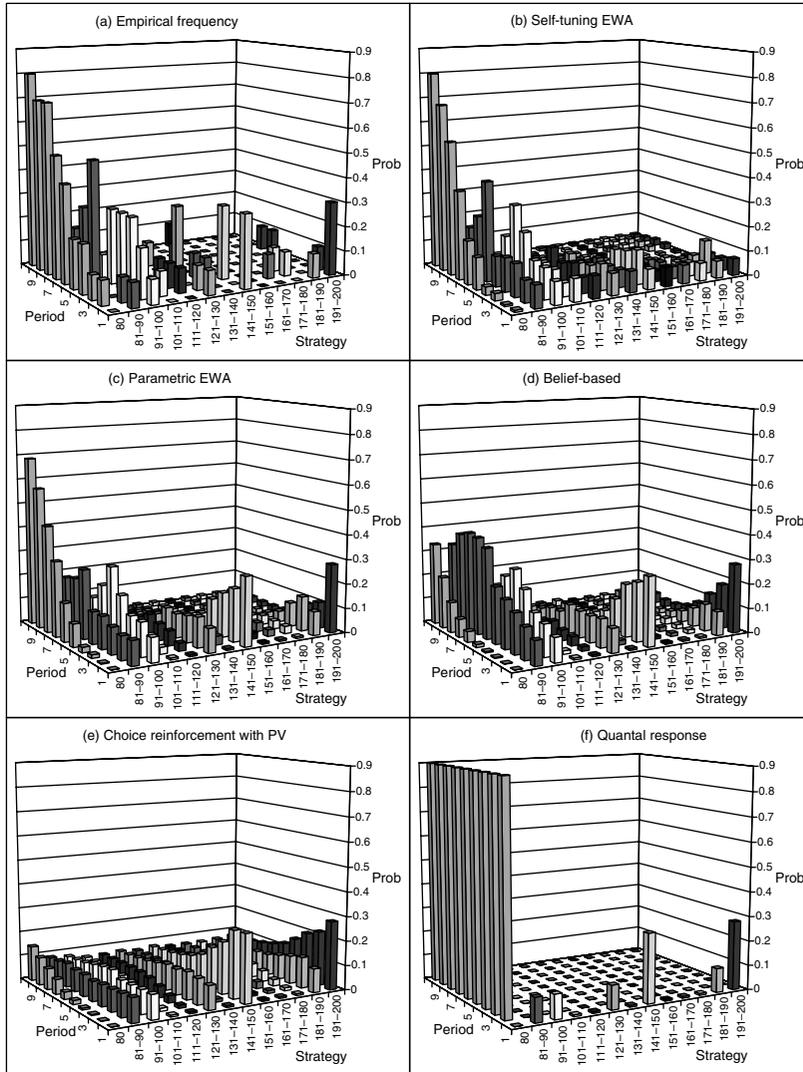


Figure 8.8 Predicted frequency for traveller's dilemma (Reward = 50)

predicted to move toward *ex-post* best responses, the model cannot explain why players learn to choose 80 so rapidly.

Economic value of learning models

Since the 1980s, the concept of economic engineering has emerged as being increasingly important from its start in the late 1970s (see Plott, 1986).

Experimentation has played an important role in this emergence (see Plott, 1997; Rassenti *et al.*, 2001). For the practice of economic engineering, it is useful to have a measure of how much value a theory or design creates. For policy purposes, increases in allocative efficiency are a sensible measure, but for judging the private value of advice to a firm or consumer other measures are more appropriate.

Camerer and Ho (2001) introduced a measure called 'economic value'. Schelling (1960, p. 98) wrote, 'a normative theory must produce strategies that are at least as good as what people can do without them'. Inspired by his definition, the economic value of a theory is how much model forecasts of behaviour of other players improve the profitability of a particular player's choices. This measure treats a theory as being like the advice service that professionals sell (for example, consultants). The value of a theory is the difference in the economic value of the client's decisions with and without the advice.

Besides being a businessperson's measure, economic value is a way of measuring the degree of disequilibrium in economic terms. Note that, in equilibrium, the economic value of a learning theory is zero or negative by definition (since players are already guessing perfectly accurately what others will do). A bad theory, which implicitly 'knows' less than the subjects themselves do about what other subjects are likely to do, will have negative economic value.

Furthermore, do not conclude, mistakenly, that if a learning theory has economic value it does not describe how people in fact learn. The economic value assumes that an objective observer uses the theory to make a forecast and best-responds to it – in our terms below, such a person is 'sophisticated'. So if the model describes accurately how adaptive players (who are *not* sophisticated) learn, it will have economic value. It *is* true, however, that a model of sophisticated players should not have economic value (since the advice it gives should already be known to the players, by definition of sophistication).

To measure economic value, we use model parameters and a player's observed experience through period t to generate model predictions about what others will do in $t + 1$. These predictions are used to compute expected payoffs from strategies, and recommend a choice with the highest expected value. We then compare the profit from making that choice in $t + 1$ (given what other players did in $t + 1$) with profit from the target player's actual choice. Economic value is a good measure because it uses the full distribution of predictions about what other players are likely to do, *and* the economic impact of those possible choices. These measures do not control for the boomerang effect of how a recommended choice would have changed future behaviour by others, but this effect is small in most games.⁴³

Data from six games are used to estimate model parameters and make recommendations in the seventh game, for each of the games separately. Table 8.5 shows the overall economic value – the percentage improvement

Table 8.5 Economic value of learning theories (percentage improvement in payoffs)

Game	self-tuning EWA(%)	parametric EWA(%)	Belief- based(%)	Reinf.- PV(%)	QRE(%)
cont'l divide	5.0	5.2	4.6	-9.4	-30.4
median action	1.5	1.5	1.2	1.3	-1.0
p-Beauty contest	49.9	40.8	26.7	-7.2	-63.5
price matching	10.3	9.8	9.4	3.4	2.7
mixed strategies	7.5	3.0	1.1	5.8	-1.8
patent race	1.7	1.2	1.3	2.9	1.2
pot games	-2.7	-1.1	-1.3	-1.9	9.9

Note: Highest economic value for each game is displayed in bold type.

(or decline) in payoffs of subjects from following a model recommendation rather than their actual choices. The highest economic value for each game is printed in bold. Most models have positive economic value. The percentage improvement is small in some games because even clairvoyant advice would not raise profits much.⁴⁴

Self-tuning EWA and EWA usually add the most value (except in pot games, where only QRE adds value). Belief learning has positive economic value in all but one game. Reinforcement learning adds the most value in patent races, but has negative economic value in three other games. (Reinforcement underestimates the rate of strategy change in continental divide and beauty contest games, and hence gives bad advice.) QRE has negative economic value in four games.

Summary

This section reports a comparison among several learning models on seven data sets. The new model is self-tuning EWA, a variant of the hybrid EWA model in which estimated parameters are replaced by functions entirely determined by data. Self-tuning EWA captures a predictable cross-game variation in parameters and hence fits better than other models when common parameters are estimated across games. A closer look at the continental divide and price-matching games shows that belief models are close to the data on average but miss other features (the asymmetry in convergence toward each of the two pure equilibria in the continental divide game, and the sharp convergence on the minimum price in price-matching). Reinforcement predicts well in co-ordination games and predicts the correct price often in price-matching (but with too little probability). However, reinforcement predicts badly in beauty contest games. It is certainly true that for explaining some features of some games, the reinforcement and belief models are adequate. But self-tuning EWA is easier to estimate (it has one parameter instead of two) and explains subtler features other models sometimes miss. It is also never fits poorly (relative to other games), which is the definition of robustness.

Sophistication and teaching

The learning models discussed in the previous section are adaptive and backward-looking: Players only respond to their own previous payoffs and knowledge about what others did. While a reasonable approximation, these models leave out two key features: adaptive players do not use information about *other* players' payoffs explicitly (though subjects in fact *do*⁴⁵); and adaptive models ignore the fact that when the same players are matched together repeatedly, their behaviour is often different than it is when they are not rematched together, generally in the direction of greater efficiency (see, for example, Van Huyck *et al.*, 1990; Andreoni and Miller, 1993; Clark and Sefton, 1999).

In this section, adaptive models are extended to include sophistication and strategic teaching in repeated games (see Stahl, 1999; and Camerer *et al.*, 2002a, for details). Sophisticated players believe that others are learning and anticipate how those others will change in deciding what to do. In learning to shoot at a moving target, for example, soldiers and fighter pilots learn to shoot *ahead*, towards where the target *will be*, rather than shoot at the target where it is when they aim. They become sophisticated.

Sophisticated players who also have strategic foresight will 'teach' – that is, they choose current actions which teach the learning players what to do, in a way that benefits the teacher in the long run. Teaching can either be mutually beneficial (trust-building in repeated games) or privately beneficial but socially costly (entry-deterrence in chain-store games). Note that sophisticated players will use information about the payoffs of others (to forecast what others will do), and will behave differently depending on how players are matched, so adding sophistication can conceivably account for the effects of information and matching that adaptive models miss.⁴⁶

Sophistication

Let us begin with myopic sophistication (no teaching). The model assumes a population mixture in which a fraction α of players are sophisticated and $1 - \alpha$ are adaptive. (It is possible to imagine a model with degrees of sophistication, as well, or learning to become sophisticated, as in Stahl, 1999.) To allow for possible over-confidence, sophisticated players think that a fraction $(1 - \alpha')$ of players are adaptive and the remaining fraction α' of players are sophisticated, like themselves.⁴⁷ Sophisticated players use the self-tuning EWA model to forecast what adaptive players will do, and choose strategies with high expected payoffs, given their forecast and their guess about what sophisticated players will do. Denoting choice probabilities by adaptive and sophisticated players by $P_i^j(a, t)$ and $P_i^j(s, t)$, attractions for sophisticates are:

$$A_i^j(s, t) = \sum_{k=1}^{m-i} [\alpha' P_{-i}^k(s, t+1) + (1 - \alpha') \times P_{-i}^k(a, t+1)] \times \pi_i(s_i^j, s_{-i}^k) \quad (8.8)$$

Note that, since the probability $P_{-i}^k(s, t + 1)$ is derived from an analogous condition for $A_i^j(s, t)$, the system of equations is recursive. Self-awareness creates a whirlpool of recursive thinking which means that QRE (and Nash equilibrium) are special cases in which all players are sophisticated and believe others are too ($\alpha = \alpha' = 1$).

An alternative structure one could study links steps of sophistication to the steps of thinking used in the first period. For example, define zero learning steps as using self-tuning EWA; one step is best-responding to zero-step learners; two steps is best-responding to choices of one-step sophisticates, and so forth. We think this model can produce results similar to the recursive one we report below, and it replaces α and α' with τ from the theory of initial conditions so reducing the entire thinking–learning–teaching model to just two parameters.

We estimate the sophisticated EWA model using data from the p -beauty contests introduced above. Table 8.6 reports results and estimates of important parameters (with bootstrapped standard errors in parentheses). For inexperienced subjects, adaptive EWA generates Cournot-like estimates ($\hat{\phi} = 0$ and $\hat{\delta} = 0.90$). Adding sophistication increases $\hat{\phi}$ and improves LL substantially both in and out of sample. The estimated fraction of sophisticated players is 24 per cent and their estimated perception $\hat{\alpha}'$ is zero (and is insignificant), showing over-confidence (as in the thinking-steps estimates from the previous section).⁴⁸

Experienced subjects are those who play a second 10-period game with a different p parameter (the multiple of the average that creates the target number).

Table 8.6 Sophisticated and adaptive learning model estimates for the p -beauty contest game

	Inexperienced subjects		Experienced subjects	
	<i>Sophisticated EWA</i>	<i>Adaptive EWA</i>	<i>Sophisticated EWA</i>	<i>Adaptive EWA</i>
ϕ	0.44 (0.05) ²	0.00 (0.00)	0.29 (0.03)	0.22 (0.02)
δ	0.78 (0.08)	0.90 (0.05)	0.67 (0.05)	0.99 (0.02)
α	0.24 (0.04)	0.00 (0.00)	0.77 (0.02)	0.00 (0.00)
α'	0.00 (0.00)	0.00 (0.00)	0.41 (0.03)	0.00 (0.00)
LL				
In sample	-2095.32	-2155.09	-1908.48	-2128.88
Out of sample	-968.24	-992.47	-710.28	-925.09

Note: ²Standard errors in parentheses.

Source: Camerer *et al.*, 2002b.

Among experienced subjects, the estimated proportion of sophisticates increases to $\hat{\alpha} = 77$ per cent. Their estimated perceptions also increase, but are still over-confident ($\hat{\alpha}' = 41$ per cent). The estimates reflect 'learning about learning': subjects who played one 10-period game came to realize that an adaptive process is occurring; and most of them anticipate that others are also learning when they play again.

Strategic teaching

Sophisticated players matched with the same players repeatedly often have an incentive to 'teach' adaptive players by choosing strategies with poor short-run payoffs that will change what adaptive players do, in a way that benefits the sophisticated player in the long run. Game theorists have shown that strategic teaching could select one of many repeated-game equilibria (teachers will teach the pattern that benefits them) and could give rise to reputation formation without the complicated apparatus of Bayesian updating of Harsanyi-style payoff types (see Fudenberg and Levine, 1989; Watson, 1993; Watson and Battigali, 1997). This section of our chapter describes a parametric model which embodies these intuitions, and tests it with experimental data. The goal is to show how the kinds of learning models described in the previous section can be extended parsimoniously to explain behaviour in more complex games which are, perhaps, of even greater economic interest than games with random matching.

Consider a finitely-repeated trust game, first studied by Camerer and Weigelt (1988). A borrower B wants to borrow money from each of a series of lenders denoted L_i ($i = 1, \dots, N$). In each period a lender makes a single lending decision (*Loan* or *No Loan*). If the lender makes a loan, the borrower either *repays* or *defaults*. The next lender in the sequence, who observed all the previous history, then makes a lending decision. The payoffs used in the experiments are shown in Table 8.7.

There are in fact two types of borrowers. As in post-Harsanyi game theory with incomplete information, types are expressed as differences in borrower payoffs which the borrowers know but the lenders do not (though the probability of a given borrower is each type is commonly known). The honest (Y) types in fact receive *more* money from repaying the loan, an experimenter's

Table 8.7 Payoffs in the borrower – lender trust game

Lender strategy	Borrower strategy	Payoffs to lender	Payoffs to borrower	
			Normal (X)	Honest (Y)
Loan	Default	–100	150	0
	Repay	40	60	60
No loan	(No choice)	10	10	10

Source: Camerer and Weigelt (1988).

way of inducing preferences like those of a person who has a social utility for being trustworthy (see Camerer, 2002, ch. 3 and references). The normal (X) types, however, earn 150 from defaulting and only 60 from repaying. If they were playing just once and wanted to earn the most money, they would default.

In the standard game-theoretic account, paying back loans in finite games arises because there is a small percentage of honest types who *always* repay. This gives normal-type borrowers an incentive to repay until close to the end, when they begin to use mixed strategies and default with increasing probability.

Whether people in fact play these sequential equilibria is important to investigate, for two reasons. First, the equilibria impose consistency between optimal behaviour by borrowers and lenders, and Bayesian updating of types by lenders (based on their knowledge and anticipation of the borrowers' strategy mixtures); whether reasoning or learning can generate this consistency is an open behavioural question (see Selten, 1978). Second, the equilibria are very sensitive to the probability of honesty (if it is too low, the reputational equilibria disappear and borrowers should always default), and also make counterintuitive comparative statics predictions that are not confirmed in experiments (for example, Neral and Ochs, 1992; Jung *et al.*, 1994).

In the experiments, subjects play many sequences of eight periods. The eight-period game is repeated to see whether equilibration occurs across many sequences of the entire game.⁴⁹ Surprisingly, the earliest experiments showed that the pattern of lending, default and reactions to default across experimental periods within a sequence is roughly in line with the equilibrium predictions. Typical patterns in the data are shown in Figure 8.9. Sequences are combined into ten-sequence blocks (denoted 'sequence') and average frequencies are reported from those blocks. Periods 1, . . . , 8 denote periods in each sequence. The figures show relative frequencies of *no loan* and *default* (conditional on a loan). Figure 8.9(a) shows that, in early sequences, lenders start by making loans in early periods (that is, there is a low frequency of no loan), but they rarely lend in periods 7–8. In later sequences they have learned always to lend in early periods and rarely in later periods. Figure 8.9(b) shows that borrowers rarely default in early periods, but usually default (conditional on getting a loan) in periods 7–8. The within-sequence pattern becomes sharper in later sequences.

The general patterns predicted by equilibrium are therefore present in the data. But given the complexity of the equilibrium, how do players approximate it? Camerer and Weigelt (1988) concluded their paper as follows:

the long period of disequilibrium behavior early in these experiments raises the important question of how people learn to play complicated games. The data could be fit to statistical learning models, though new experiments or new models might be needed to explain learning adequately. (pp. 27–8)

The teaching model is a 'new model' of the sort Camerer and Weigelt had in mind. It is a boundedly rational model of reputation formation in which

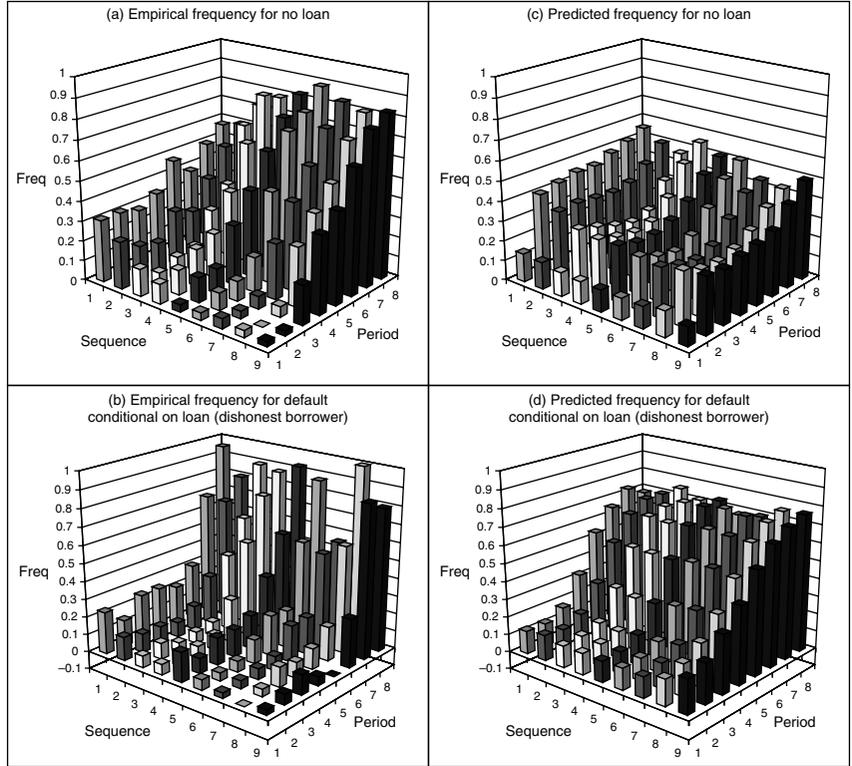


Figure 8.9 Empirical and predicted frequencies for borrower-lender trust game

the lenders learn whether to lend or not. They do not update borrowers' types and do not anticipate borrowers' future behaviour (as in equilibrium models); they just learn.

In the teaching model, a proportion of borrowers are sophisticated and teach; the rest are adaptive and learn from experience but have no strategic foresight. The teachers choose strategies that are expected (given their beliefs about how borrowers will react to their teaching) to give the highest long-run payoffs in the remaining periods.

A sophisticated teaching borrower's attractions for sequence k after period t are specified as follows ($j \in \{repay, default\}$ is the borrower's set of strategies):

$$A_B^i(s, k, t) = \sum_{j'=Loan}^{No Loan} P_L^{j'}(a, k, t+1) \cdot \pi_B(j, j') + \max_{j_{t+1}} \left\{ \sum_{v=t+2}^T \sum_{j'=Loan}^{No Loan} \hat{P}_L^{j'}(a, k, v | j_{v-1} \in J_{t+1}) \cdot \pi_B(j_v \in J_{t+1}, j') \right\}$$

The set J_{t+1} specifies a possible path of future actions by the sophisticated borrower from round $t + 1$ until end of the game sequence. That is, $J_{t+1} = \{j_{t+1}, j_{t+2}, \dots, j_{T-1}, j_T\}$ and $j_{t+1} = j$.⁵⁰ The expressions $\hat{P}_L^j(a, k, v | j_{v-1})$ are the overall probabilities of either getting a loan or not in the future period v , which depends on what happened in the past (which the teacher anticipates).⁵¹ $P_B^j(s, k, t + 1)$ is derived from $A_B^j(s, k, t)$ using a logit rule.

The updating equations for adaptive players are the same as those used in self-tuning EWAs, with two twists. First, since lenders who play in later periods know what has happened earlier in a sequence, we assume that they learnt from the experience they witnessed as though it had happened to them.⁵² Second, a lender who is about to make a decision in Period 5 of Sequence 17, for example, has *two* relevant sources of experience on which to draw – the behaviour seen in Periods 1–4 in Sequence 17, *and* the behaviour seen in the Period 5s of the previous sequences (1–16). Since *both* kinds of experience could influence the lender's current decision, we include both, using a two-step procedure. After Period 4 of Sequence 17, for example, attractions for lending and not lending are first updated, based on the Period 4 experience. Then attractions are partially updated (using a degree of updating parameter σ) based on the experience in Period 5 of the previous sequences.⁵³ The parameter σ is a measure of the strength of 'peripheral vision' – glancing back at the 'future' Period 5s from previous sequences to help guess what lies ahead.

Of course, it is well known that repeated-game behaviour can arise in finite-horizon games when there are a small number of 'unusual' types (who act as though the horizon is unlimited), which creates an incentive for rational players to behave as if the horizon is unlimited until near the end (for example, Kreps and Wilson, 1982). But specifying why some types are irrational, and how many there are, makes this interpretation difficult to test. In the teaching approach, which 'unusual' type the teacher pretends to be arises endogenously from the payoff structure: they are Stackelberg types, who play the strategy they would choose if they could commit to it. For example, in trust games, they would like to commit to repaying; but in entry-deterrence, they would like to commit to fighting entry.

The model is estimated using repeated game trust data from Camerer and Weigelt (1988). In Camerer *et al.* (2002), we used parametric EWA to model behaviour in trust games. This model allows two different sets of EWA parameters for lenders and borrowers. In this chapter we use self-tuning EWA to model lenders and adaptive borrowers so the model has fewer parameters.⁵⁴ Maximum likelihood estimation is used to estimate parameters on 70 per cent of the sequences in each experimental session, then behaviour in the holdout sample of 30 per cent of the sequences is forecast using the estimated parameters.

As a benchmark alternative to the teaching model, we estimated an agent-based version of QRE suitable for extensive-form games (see McKelvey and

Palfrey, 1998). Agent-based QRE is a good benchmark because it incorporates the key features of repeated-game equilibrium – strategic foresight, accurate expectations about actions of other players, and Bayesian updating – but assumes stochastic best-response. We use an agent-based form in which players choose a distribution of strategies at each node, rather than using a distribution over all history-dependent strategies. We implement agent QRE (AQRE) with four parameters – different λ s for lenders, honest borrowers and normal borrowers, and a fraction θ , the percentage of players with normal-type payoffs who are thought to act as if they are honest (reflecting a ‘homemade prior’ which can differ from the prior induced by the experimental design⁵⁵). (Standard equilibrium concepts are a special case of this model when λ s are large and $\theta = 0$, and fit much worse than does AQRE). The implementation in our 2002 paper is itself a small contribution since it is quite complex to estimate AQRE in these games.

The models are estimated separately on each of the eight sessions to gauge cross-session stability. Since pooling sessions yields similar fits and parameter values, we report only those pooled results in Table 8.8 (excluding the λ values). The interesting parameters for sophisticated borrowers are estimated to be $\hat{\alpha} = 0.89$ and $\hat{\sigma} = 0.93$, which means that most subjects are classified as teachers and they put a lot of weight on previous sequences. The teaching model fits in-sample and predicts better out-of-sample than AQRE by a modest margin (and does better in six out of eight individual experimental sessions), predicting about 75 per cent of the choices correctly. The AQRE fits reasonably well too (72 per cent correct) but the estimated ‘homemade prior’ θ is 0.91, which is absurdly high. (Earlier studies estimated numbers around 0.1–0.2.) The model basically fits best by assuming that *all* borrowers simply prefer to repay loans. This assumption fits most of the data but it mistakes

Table 8.8 Model parameters and fit in repeated trust games

	Statistic	Model	
		Self-tuning EWA+ teaching	Agent QRE
In-sample	Hit rate (%)	76.5%	73.9%
Calibration (n = 5757)	log-likelihood	–2975	–3131
Out-of-sample	Hit rate (%)	75.8%	72.3%
Validation (n = 2894)	log-likelihood	–1468	–1544
Parameters		Estimates	
Cross-sequence learning	σ	0.93	–
Percentage of teachers	α	0.89	–
Homemade prior p(honest)	θ	–	0.91

teaching for a pure repayment preference. As a result, it does not predict the sharp upturn in defaults in Periods 7–8, which the teaching model does.⁵⁶

Figure 8.9(c)–(d) show average predicted probabilities from the teaching model for the no-loan and conditional default rates. No-loan frequencies are predicted to start low and rise across periods, as they in fact do, though the model underpredicts the no-loan rate in general. The model predicts the increase in default rate across periods reasonably well, except for underpredicting default in the last period.

The teaching approach is a boundedly-rational alternative to type-based equilibrium models of reputation formation.⁵⁷ It has always seemed improbable that players are capable of the delicate balance of reasoning required to implement the type-based models, unless they learn the equilibrium through some adaptive process. The teaching model is one parametric model of that adaptive process. It retains the core idea in the theory of repeated games – namely, strategic foresight – and consequently respects the fact that matching protocols matter. And since the key behavioural parameters (α and σ) appear to be near 1, restricting attention to these values should make the model workable for doing theory.

Summary

In this section we introduced the possibility that players can be sophisticated – that is, they believe others are learning. (In future work, it would be interesting to link steps of iterated thinking, as in the first section, to steps of sophisticated thinking.) Sophistication links learning theories to equilibrium ones if sophisticated players are self-aware. Adding sophistication also improves the fit of data from repeated beauty-contest games. Interestingly, the proportion of estimated sophisticates is around a quarter when subjects are inexperienced, but rises to around three-quarters when they play an entire 10-period game for a second time, as if the subjects learn about learning. Sophisticated players who know they will be rematched repeatedly may have an incentive to ‘teach’, which provides a boundedly rational theory of reputation formation. We apply this model to data on repeated trust games. The model adds only two behavioural parameters, representing the fraction of teachers and how much ‘peripheral vision’ learners have (and some nuisance λ parameters), and predicts substantially better than a quantal response version of equilibrium.

Conclusion

In the introduction we stated that the research programme in behavioural game theory has three goals: (i) to create a theory of one-shot or first-period play using an index of bounded rationality measuring steps of thinking; (ii) to predict features of equilibration paths when games are repeated; and (iii) to explain why players behave differently when matched together repeatedly.⁵⁸

The models described in this chapter illustrate ways to understand these three phenomena. There are many alternative models (especially of learning). The models described here are just some examples of the style in which ideas can be expressed, and how data are used to test and modify them.

Keep in mind that the goal is *not* to list deviations from the Nash equilibrium and stop there. Deviations are just hints. The goal is to develop alternative models which are precise, general and disciplined by data – that is, some day game theory might be taught beginning with behavioural models and *ending* with analytical concepts such as equilibrium and its many refinements. All such models deserve to be called ‘game theory’, except that these models are *behavioural* – rooted in psychological regularity and sharpened by having to explain data – while analytical models are simply useful fictional answers to questions about how players of varying degrees of hypothetical rationality behave.

The thinking-steps model posits a Poisson distribution (with mean τ) of numbers of thinking steps, along with decision rules for what players using each number of steps will do. Studies with simple matrix games, beauty contests (unreported), mixed games, and entry games all show that values of τ around 1.5 can fit data reasonably well (and never worse than the Nash equilibrium). The model is easy to use because players can be assumed to best-respond and the model usually makes realistic probabilistic predictions because the mixture of thinking steps types creates a population mixture of responses. The surprise is that the same model, which is tailor-made to produce spikes in dominance-solvable games, can also fit data from games with pure and mixed equilibria using roughly the same τ .

The second section compared several adaptive learning models. For explaining simple trends in equilibration, many of these models are close substitutes. However, it is useful to focus on where models fail if the goal is to improve them. The EWA hybrid was created to include the psychological intuitions behind both reinforcement learning (received payoffs receive more weight than forgone payoffs) and belief learning (both types of payoff receive equal weight). If both intuitions were compelling enough for people to want to compare them statistically, then a model that had both intuitions in it should be better still (and generally, it is). Self-tuning EWA uses one parameter (λ) and substitutes functions for parameters. The major surprise here is that functions such as the change-detector $\phi_i(t)$ can reproduce differences across games in which parameter values fit best. This means that the model can be applied to brand-new games (when coupled with a thinking-steps theory of initial conditions) without having to make a prior judgement about which parameter values are reasonable, and without positing game-specific strategies. The interaction of learning and game structure creates reasonable parameter values automatically.

In the third section we extended the adaptive learning models to include sophisticated players who believe that others are learning. Sophistication

improves fit in the beauty contest game data. (Experienced subjects seem to have ‘learned about learning’ because the percentage of apparently sophisticated players is higher and convergence is faster.) Sophisticated players who realize they are matched with others repeatedly often have an incentive to ‘teach’, as in the theory of repeated games. Adding two parameters to adaptive learning was used to model learning and teaching in finitely-repeated trust games. While trustworthy behaviour early in these games is known to be rationalizable by Bayesian – Nash models with ‘unusual’ types, the teaching model creates unusual types from scratch. Teaching also fits and predicts better than more forgiving quantal response forms of the Bayesian – Nash type-based model. The surprise here is that the logic of mutual consistency and type updating is not needed to produce accurate predictions in finitely-repeated games with incomplete information.

Potential applications

A crucial question is whether behavioural game theory can help to explain naturally-occurring phenomena. We conclude the chapter with some speculations about the sorts of phenomena precise models of limited thinking, learning and teaching could illuminate.

Bubbles

Limited iterated thinking is potentially important because, as Keynes and many others have pointed out, it is not always optimal to behave rationally if you believe others are not. For example, prices of assets should equal their fundamental or intrinsic value if rationality is common knowledge (Tirole, 1985). But when the belief that others might be irrational arises, bubbles can too. Besides historical examples such as Dutch tulip bulbs and the \$5 trillion tech-stock bubble in the 1990s, experiments have shown such bubbles even in environments in which the asset’s fundamental value is controlled and commonly known.⁵⁹

Speculation and competition neglect

The ‘Groucho Marx theorem’ says that traders who are risk-averse should not speculate by trading with each other even if they have private information (since the only person who will trade with you may be better-informed). But this theorem rests on unrealistic assumptions of common knowledge of rationality and is violated constantly by massive speculative trading volume and other kinds of betting, as well as in experiments.⁶⁰

Players who do limited iterated thinking, or believe others are not as smart as themselves, will neglect competition in business entry (see Camerer and Lovo, 1999; Huberman and Rubinstein, 2000). Competition neglect may partly explain why the failure rate of new businesses is so high. Managerial hubris, over-confidence and self-serving biases correlated with costly delay

and labour strikes in the lab (Babcock *et al.*, 1995) and in the field (Babcock and Loewenstein, 1997) can also be interpreted as players not believing that others always behave rationally.

Incentives

In a thorough review of empirical evidence on incentive contracts in organizations, Prendergast (1999) notes that workers typically react to simple incentives, as standard models predict. However, firms do not usually implement complex contracts, which *should* elicit greater effort and improve efficiency. Perhaps the firms' reluctance to bet on rational responses by workers is evidence of limited iterated thinking.

Macroeconomics

Woodford (2001) notes that in Phelps–Lucas 'islands' models, nominal shocks can have real effects, but their predicted persistence is too short compared to effects in data. He shows that imperfect information about *higher-order* nominal gross domestic product (GDP) estimates – beliefs about beliefs, and higher-order iterations – can cause longer persistence which matches the data. However, Svensson (2001) notes that iterated beliefs are probably constrained by computational capacity. If people have a projection bias, their beliefs about what others believe will be too much like their own, which undermines Woodford's case. On the other hand, in the thinking-steps model, players' beliefs are not mutually consistent so there is higher-order belief inconsistency which can explain longer persistence. In either case, knowing precisely how iterated beliefs work could help to inform a central issue in macroeconomics – persistence of the real effects of nominal shocks.

Learning

Other phenomena are evidence of a process of equilibration or learning. For example, institutions for matching medical residents and medical schools, and analogous matching in college sororities and college bowl games, developed over decades and often 'unravel' so that high-quality matches occur before some agreed-upon date (Roth and Xing, 1994). Bidders in eBay auctions learn to bid late to hide their information about an object's common value (Bajari and Hortacsu, 2003). Consumers learn over time what products they like (Ho and Chong, 2003). Learning in financial markets can generate excess volatility and returns predictability, which are otherwise anomalous in rational expectations models (Timmerman, 1993). We are currently studying evolution of products in a high-uncertainty environment (electronics equipment) for which thinking-steps and learning models are proving useful.

Teaching

Teaching in repeated games may prove to be the most potentially useful tool for economics, because it is essentially an account of how bounded rationality can give rise to some features of repeated-game behaviour, where standard theory has been applied widely. The teaching model could be applied to repeated contracting, employment relationships, alliances among firms, industrial organization problems (such as pricing games among perennial rivals, and entry deterrence) and macroeconomic models of policy-maker inflation-setting. For example, the teaching model could be applied to the Kydland – Prescott model of commitment in which the public learns about inflation from past history (using the self-tuning EWA rule described in the Appendix) and unemployment is determined by an expectational Phillips curve. Since policy-makers face a temptation to choose surprisingly high inflation to lower unemployment, they can either act myopically or ‘teach’ the public to expect low inflation which is Pareto-optimal in the long-run (see Sargent, 1999). A model similar to self-tuning EWA applied to hyperinflation is Marcet and Nicolini (2003).

Appendix: Thinking models applied to mixed games and entry games

Games with mixed equilibria

A good model of thinking steps should be able both to account for deviations from the Nash equilibrium (as in the games above), *and* reproduce the successes of the Nash equilibrium. A domain in which the Nash equilibrium does a surprisingly good job is in games with unique mixed equilibria. It is hard to beat the Nash equilibrium in these games because (as we shall see) the correlation with data is in fact very good (around 0.9) so there is little room for improvement. Instead, the challenge is to see how well a thinking-steps model which bears little resemblance to the algebraic logic of equilibrium mixing can approximate behaviour in these games.

Early tests in the 1960s and 1970s (mostly by psychologists) appeared to reject the Nash equilibrium as a description of play in mixed games. As others have noted (for example, Binmore *et al.*, 2001), these experiments were incomplete in important dimensions and hence inconclusive. Financial incentives were very low or absent; subjects typically did not play other human subjects (and often were deceived about playing other people, or were instructed only vaguely about how their computer opponents played); and pairs were often matched repeatedly so that (perceived) detection of temporal patterns permitted subjects to choose non-equilibrium strategies. Under conditions ideal for equilibration, however, convergence was rapid and sharp. Kaufman and Becker (1961), for example, asked subjects to specify mixtures and told them that a computer program would then choose a mixture to minimize the subjects’ earnings. Subjects could maximize their possible gains by choosing the Nash mixture. After playing five games, more than half learned to do so. More recent experiments are also surprisingly supportive of the Nash equilibrium (see Binmore *et al.*, 2001; Camerer, 2002, ch. 2). The data are supportive in two senses: (i) equilibrium predictions and actual frequencies are closely correlated, when taken as a whole (for example, strategies predicted to be more likely are almost always played more often);

and (ii) it is hard to imagine any parsimonious theory that can explain the modest deviations.

We applied a version of the thinking model in which K -step thinkers think all others are using $K - 1$ steps along with best response to see whether it could produce predictions as accurate as Nash in games with mixed equilibria. This model is extremely easy to use (just start with step-zero mixtures and compute best responses iteratively). Furthermore, it creates natural ‘purification’: players using different thinking steps usually choose pure strategies, but the Poisson distribution of steps generates a mixture of responses, and hence a probabilistic prediction.

Model predictions are compared with data from fifteen games with unique mixed equilibria, reported in Camerer (2002, ch. 2).⁶¹ These games are not a random or exhaustive sample of recent research but there are enough observations that we are confident the basic conclusion will be overturned by adding more studies. Note that we use data from *all* the periods of these games rather than the first period only. (In most cases, the first-period data are rarely reported, and there is usually little trend over time in the data.)

Figure 8.A1 plots actual frequencies on the ordinate (y) axis against either mixed-strategy equilibrium predictions or thinking-steps predictions on the abscissa (x) axis. Each data point in Figure 8.A1 represents a single strategy from a different game (pooling across all periods to reduce sampling error).⁶²

In Figure 8.A1, the value of τ is common across games (1.46) and minimizes mean squared deviations between predicted and actual frequencies. When values are estimated separately for each game to minimize mean squared deviations, the values across the games (in the order they are listed above) are 0.3, 0.3, 0.3, 2.2, 2.5, 0.1, 1.8, 2.3, 2.9, 2.7, 0.5, 0.8, 1.6, 1.5, 1.9. The lower values occur in games where the actual

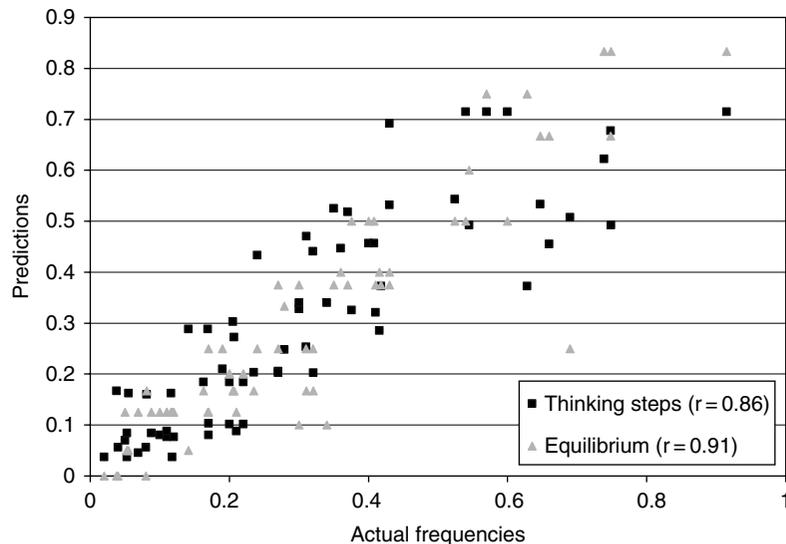


Figure 8.A1 Fit of data to equilibrium and thinking-steps model (common $\tau = 1.5$) in mixed-equilibrium games

mixtures are close to equal across strategies, so that a distribution with $\tau = 0$ fits well. When there are dominated strategies, which are usually rarely played, much higher τ values are needed, since low τ generates a lot of random play and frequent dominance violation. The simple arithmetical average across the fifteen games is 1.45, which is very close to the best-fitting common $\tau = 1.46$. Figure 8.A1 shows two regularities: both thinking-steps (circles in the plot) and equilibrium predictions (triangles) have very high correlations (above 0.8) with the data, though there is substantial scatter, especially at high probabilities. (Keep in mind that sampling error means there is an upper bound on how well any model could fit – even the true model that generated the data.) The square root of the mean squared deviation is around 0.10 for both models.

While the thinking-steps model with common τ is a little less accurate than the Nash equilibrium (the game-specific model is more accurate), the key point is that the same model that can explain Nash deviations in dominance-solvable games and matrix games fits almost as well with a value of τ close to those estimated in other games.

Table 8.A1 shows a concrete example of how the thinking model is able to approximate mixture probabilities. The game is Mookerjee and Sopher's (1997) 4×4 game. Payoffs are shown as wins (+) or losses (–) ($2/3 +$ means a $2/3$ chance of winning) in the upper left cells. The rightmost columns show the probabilities with which row players using different thinking steps choose each of the four row strategies. To narrate a little, zero-step players randomize (each is played with probability of 0.25); one-step players best-respond to a random column choice and choose row strategy 3 with probability of 1, and so on. First notice that the weakly dominated strategy (4) is only chosen by a quarter of zero-step players (since it is never the best response against players who randomize), which generates a small amount of choice that matches the data. Notice also that the best responses tend to lurch around as thinking-steps changes. When these are averaged over thinking-steps frequencies, a population mixture results. Furthermore, one of the quirkiest features of mixed equilibrium is that one player's mixture depends only on the other player's payoffs. This effect also occurs in the thinking steps models, because a K -step row player's payoffs affect row's best responses, which then affect a $K + 1$ -step column player's best response. So one player's payoffs affect the other's strategies indirectly. Table 8.A1 also shows the MSE (mixed equilibrium) prediction, the data frequencies and overall frequencies from the thinking-steps model when $\tau = 2.2$. The model fits the data more closely than MSE for row players (it accounts for underplay of row 2 strategy and overplay of strategy 3) and is as accurate as MSE for column players. As noted in the text, the point is not that the thinking-steps model outpredicts MSE – it cannot, because MSE has such a high correlation – but simply that the model that explains behaviour in dominance-solvable, matrix and entry games also generates mixtures of players playing near-pure strategies that are close to outcomes in mixed games.

Market entry games

Analysis of the simple entry game described in the text proceeds as follows. Step 0s randomize so $f(0)/2$ level 0s enter. Define the relative proportion of entry after accounting up through level k as $N(k)$. Define a Boolean function $B(X) = 1$ if X true, $B(X) = 0$ if X false. Level 1s enter iff $1/2 < c$. Therefore, total entry 'after' Level 1 types are accounted for is $N(1) = f(0)/2 + B[N(0)/(f(0)/2) > c]f(1)$. Total entry after Level k type is therefore $N(k) = f(0)/2 + \sum_{n=1}^k f(n)B[N(n-1)/(\sum_{m=1}^k f(m)) > c]$. A given c and τ then generates a sequence of cumulated entry rates which asymptotes as k grows

Table 8.A1 How thinking steps fits mixed-game data

Row strategies	Strategies 1-4				Column thinking steps 0-5					MSE	Data	Thinking Model	
	1	2	3	4	Step 0	1	2	3	4				5
1	+	-	-	+	0.25	0	0.5	1	0	0	0.375	0.32	0.37
2	-	-	+	+	0.25	0	0	0	1	0	0.25	0.17	0.14
3	-	+	(2/3)+	(2/3)+	0.25	1	0.5	0	0	1	0.375	0.43	0.46
4	-	-	(1/3)+	+	0.25	0	0	0	0	0	0	0.08	0.03
Steps													
0	0.25	0.25	0.25	0.25									
1	0.5	0.5	0	0									
2	1	0	0	0									
3	0	0	1	0									
4	0	0.5	0.5	0									
5	0.5	0.5	0	0									
6	1	0	0	0									
7	0	0	1	0									
MSE	0.375	0.25	0.375	0									
Data	0.38	0.31	0.27	0.04									
Thinking	0.46	0.23	0.28	0.03									

Source: (Mookerjee and Sopher 4 x 4 game)

large. Define a function $N(all|\tau)(c)$ as the overall rate of entry, given τ , for various capacities of c .

The data reported in the text come from experiments by Sundali *et al.* (1995) and Seale and Rapoport (2000). Their game is not precisely the same as the one analyzed because, in their game, entrants earn $1 + 2(c - e)$ (where e is the number of entrants) and non-entrants earn 1. They used twenty subjects with odd values of $c(1, 3, \dots, 19)$. To compute entry rates reported in Figure 8.A1 we averaged entry for adjacent c values (that is, averaging 1 and 3 yields a figure for $c = 2$ matching $c = 0.1$, averaging 3 and 5 yields a figure for $c = 4$ corresponding to $c = 0.2$ and so on). Obviously, the analysis and data are not perfectly matched, but we conjecture that the thinking-steps model can still match data closely and reproduce the three experimental regularities described in the text; whether this is true is the subject of ongoing research.

Notes

- 1 Our models are related to important concepts such as rationalizability, which weakens the mutual consistency requirement, and behaviour of finite automata. The difference is that we work with simple parametric forms and concentrate on fitting them to data.
- 2 In his thesis proposing a concept of equilibrium, Nash himself suggested that equilibrium might arise from some ‘mass action’ that adapted over time. Taking up Nash’s implicit suggestion, later analyses filled in details of where evolutionary dynamics lead (see Weibull, 1995; Mailath, 1998).
- 3 While great triumphs of economic theory come from parameter-free models (for example, Nash equilibrium), relying on a small number of free parameters is more typical in economic modelling. For example, nothing in the theory of intertemporal choice pins a discount factor δ to a specific value. But if a wide range of phenomena are consistent with a value such as 0.95, for example, then as economists we are comfortable working with such a value despite the fact that it does not emerge from axioms or deeper principles.
- 4 Lucas (1986) makes a similar point in macroeconomic models. Rational expectations often yield indeterminacy, whereas adaptive expectations pin down a dynamic path. Lucas writes: ‘The issue involves a question concerning how collections of people behave in a specific situation. Economic theory does not resolve the question . . . It is hard to see what can advance the discussion short of assembling a collection of people, putting them in the situation of interest, and observing what they do’ (p. S421).
- 5 Quantal response equilibrium (QRE), a statistical generalization of Nash, almost always explains the direction of deviations from Nash and should replace Nash as the static benchmark to which other models are routinely compared. Stahl and Wilson (1995), Capra (1999) and Goeree and Holt (2003) have models of limited thinking in one-shot games which are similar to ours. Jehiel (2002) proposes a concept of limited foresight in games in which analogy is used to truncate complex games. There are many learning models. Self-tuning EWA generalizes some of them (though reinforcement with payoff variability adjustment is different; see Erev *et al.*, 1999). Other approaches include rule learning (Stahl, 1996, 2000), and earlier artificial intelligence (AI) tools such as genetic algorithms or genetic programming to ‘breed’ rules (see Jehiel, fourth coming). Finally, there are no alternative models of strategic teaching that we know of but this is an important area others should examine.

- 6 Note the timing convention – attractions are defined *before* a period of play; so the initial attractions $A_i^1(0)$ determine choices in Period 1, and so on.
- 7 This concept was first studied by Stahl and Wilson (1995) and Nagel (1995), and later by Ho, *et al.* (1998). A 1993 working paper by Debra Holt that also pioneered this approach was published in 1999. See also also Sonsino *et al.* (2000).
- 8 Embedded sentences are those in which subject–object clauses are separated by other subject–object clauses. A classic example is: ‘The mouse that the cat that the dog chased bit ran away’. To answer the question ‘Who got bitten?’ the reader must keep in mind ‘the mouse’ while processing the fact that the cat was chased by the dog. Limited working memory leads to frequent mistakes in recalling the contents of such sentences or answering questions about them (Christiansen and Chater, 1999). This notation makes it easier: ‘The mouse that (the cat that [the dog chased] bit) ran away’.
- 9 A natural competitor to the thinking-steps model for explaining one-shot games is quantal response equilibrium (QRE) (see McKelvey and Palfrey, 1995, 1998; Goeree and Holt, 1999a). Weiszacker (2003) suggests an asymmetric version equivalent to a thinking-steps model in which one type thinks others are more random than s/he is. More cognitive alternatives are the theory of thinking trees of Capra (1999) and the theory of ‘noisy introspection’ of Goeree and Holt (2004). In Capra’s model, players introspect until their choices match those of players whose choices they anticipate. In Goeree and Holt’s theory, players use an iterated quantal response function with a response sensitivity parameter equal to λ/t^n where n is the discrete iteration step. When t is very large, their model corresponds to one in which all players do one-step and think others do zero. When $t = 1$ the model is QRE. All these models generate unimodal distributions so they need to be expanded to accommodate heterogeneity. Further work should try to distinguish different models, or investigate whether they are similar enough to be close modelling substitutes.
- 10 The data are 48 subjects playing 12 symmetric 3×3 games (Stahl and Wilson, 1995); 187 subjects playing 8 2×2 asymmetric matrix games (Cooper and Van Huyck, 2003); and 36 subjects playing 13 asymmetric games ranging from 2×2 to 4×2 (Costa-Gomes *et al.*, 2001).
- 11 While the common- τ models have one more free parameter than QRE, any reasonable information criterion penalizing the LL would select the thinking model.
- 12 When λ is set to 100 the τ estimates become very regular, around 2, which suggests that the variation in estimates is caused by poor identification in these games.
- 13 The differences in LL across game-specific and common τ are 0.5, 49.1 and 9.4. These are marginally significant (apart from Cooper–Van Huyck).
- 14 Take the over-confidence $K - 1$ model. The 70 per cent frequency of X choices without belief-prompting is consistent with this model if $f(0|\tau)/2 + f(1|\tau) = 0.70$, which is satisfied most closely when $\tau = 0.55$. If belief-prompting moves all thinking up one step, then the former zero-steppers will choose X and all others choose Y . When $\tau = 0.55$ the fraction of level zeros is 29 per cent, so this simple model predicts 29 per cent choice of X after belief-prompting, close to the 30 per cent that is observed.
- 15 Information measures are crucial to resolving the question of whether offers that are close to equal splits are equilibrium offers which reflect fairness concerns, or reflect at limited look-ahead and heuristic reasoning. The answer is both (see Camerer *et al.*, 1993; Johnson *et al.*, 2002). In the Costa-Gomes (2001) study, two different decision rules always led to the same choices in their games, but required

- different look-up patterns. The look-up data were able therefore to classify players according to decision rules more conclusively than could choices alone.
- 16 To name only a few examples, see Camerer (1987) in markets for risky assets (partial adjustment models); Smith *et al.* (1988) in finitely-lived asset markets (Walrasian excess demand); McAllister (1991) in co-ordination games (reinforcement); Camerer and Weigelt (1993) in stochastically-lived assets (entrepreneurial stockpiling); Roth and Erev (1995) in bargaining and best-shot public goods games (reinforcement learning); Ho and Weigelt (1996) in extensive-form co-ordination games (reinforcement and belief learning); and Camerer and Cachon (1996) in co-ordination games (Cournot dynamics).
 - 17 The parameter ρ is different from the one used in our original paper and is equivalently called 'kappa' in our recent working papers.
 - 18 We have also applied slight variants of the model to extensive-form centipede games (Ho *et al.*, 1999), extensive-form signalling games (Anderson and Camerer, 2001); and bilateral call markets with private values (Camerer *et al.*, 2002). Adapting the model to extensive forms and incomplete information requires making very slight further assumptions, but the three studies we have done show that the same basic framework can be used successfully. In extensive-form games the model needs to estimate forgone payoffs for branches of the game tree that are not reached. In the papers listed above, we do this using historical information, some average of the minimum and maximum payoffs (in signalling and centipede games), or 'clairvoyance' (the other player's stated conditional response to the unchosen path in the tree, in centipede games). Any of these assumptions fit equally well. In the bilateral call markets we 'spill over' reinforcement from one private value to neighbouring private values (this kind of generalization is assumed in belief learning), but the strength of spillover depends on how close the private values are; see also Roth and Erev (1995).
 - 19 Other models aim to explain choices aggregated at some level. Of course, models of this sort can sometimes be useful. But our view is that a parsimonious model that can explain very fine-grained data can probably explain aggregated data well too, while the opposite may not be true, so the harder challenge is to explain fine-grained data.
 - 20 See also Cheung and Friedman, 1997, pp. 54–5; Fudenberg and Levine, 1998, pp. 184–5; and Ed Hopkins, 2002.
 - 21 See Bush and Mosteller, 1955; Harley, 1981; Cross, 1983; Arthur, 1991; McAllister, 1991; Roth and Erev, 1995; Erev and Roth, 1998.
 - 22 When updated fictitious play beliefs are used to update the expected payoffs of strategies, precisely the same updating is achieved by reinforcing all strategies by their payoffs (whether received or forgone). The beliefs themselves are an epiphenomenon that disappear when the updating equation is written expected payoffs rather than beliefs.
 - 23 Weber (fourth coming) also finds that players' choices change over time, as if they are learning, even with *no* feedback about what others have done (that is, even when they choose strategies repeatedly with no feedback). He suggests that players are using their own previous choices as a kind of pseudo-experience.
 - 24 Taken seriously, reinforcement models also predict that learning paths will look the same whether players know their full payoff matrix or not. This prediction is rejected in all the studies that have tested it – for example, Mookerjee and Sopher, 1994; Rapoport and Erev, 1998; Van Huyck *et al.*, 2001.

- 25 In mixed games, no model improves much on the Nash equilibrium (and often does not improve on quantal response equilibrium at all), and parameter identification is poor; see Salmon, 2001).
- 26 Typically, we penalize in-sample likelihood functions using the Akaike and Bayesian information criteria, which subtract a penalty of one, or $\log(n)$, times the number of degrees of freedom from the maximized likelihood. More persuasively, we rely mostly on out-of-sample forecasts which will be *less* accurate if a more complex model simply appears to fit better because it overfits in sample.
- 27 Note that, if your statistical objective is to maximize hit rate, λ does not matter and so self-tuning EWA is a zero-parameter theory given initial conditions.
- 28 Note that, if there is more than one other player, and the distinct choices by different other players matter to player i , then the vector is an $n - 1$ -dimensional matrix if there are n players.
- 29 In a naturally-occurring application where the equilibrium structure is not known to the modeller, it is easy to create some statistic that proxies for W , such as the number of strategies that have been chosen more than 25 per cent of the time in the previous eight periods. Such a rule will often overestimate W in empirical application, biasing $\phi_i(t)$ upwards compared to the value created by knowing W exactly. Since letting $\phi_i(t)$ dip too low, and discarding history irretrievably is a bigger mistake than keeping too much history, such an adjustment typically should not be a big mistake.
- 30 If one is uncomfortable assuming subjects act as if they know W , you can easily replace W by some function of the variability of others' choices to proxy for W .
- 31 Formally, $\rho_i(t) = 2 \times \{\sum_{k=1}^{m_i} f_i^{(k)}(t) \times \frac{m_i - k}{m_i - 1}\}$ where $f_i^k(t)$ are ranked from the lowest to the highest.
- 32 In our recent work we have switched to a model in which δ is 1 for strategies that are better (or equally good) responses to the current strategy, and zero for worse responses. This in fact fits a little worse than $\delta_i(t) = \phi_i(t)/W$, but has a sensible interpretation in terms of direction learning and dual-system dynamics. We also fix $\rho = 0$ in recent work, since allowing the flexible $\rho_i(t)$ function adds little to fit.
- 33 The initial parameter values are $\phi_i(0) = \rho_i(0) = 0.5$ and $\delta_i(0) = \phi_i(0)/W$. These initial values are averaged with period-specific values determined by the functions, weighting the initial value by $\frac{1}{t}$ and the functional value by $\frac{t-1}{t}$.
- 34 For example, suppose half the players choose *ABABAB* . . . in a game, and half choose *BABAB* Then a model which says that *A* is chosen half the time will fit perfectly in every period, even though it does not predict the transition from *A* to *B* and vice versa accurately at all.
- 35 The parameter λ seems to vary reliably across games in ways that are not easily adjusted for by rescaling payoffs to common inflation-adjusted currency levels.
- 36 More details are in Ho *et al.*, 2001, and corresponding graphs for *all* games can be seen at <http://www.bschool.nus.edu.sg/depart/mk/fbacjk/ewalite/ewalite.htm>.
- 37 Subjects were 196 undergraduate students in computer science and engineering in Singapore. Each group played ten times together twice, with different values of p in the two 10-period sequences. (One sequence used $p > 1$ and is not included.) We analyse a subsample of their data with $p = 0.7$ and 0.9 , from groups of size 7. This subsample combines groups in a 'low experience' condition (the game is the first of two they play) and a 'high experience' condition (the game is the second of two, following a game with $p > 1$).
- 38 Reinforcement can be speeded up in such games by reinforcing unchosen strategies in some way, as Roth and Erev (1995) did in market games, which is why EWA and belief learning do better.

- 39 Their experiment used ten cohorts of seven subjects each, playing for fifteen periods. At the end of each period subjects learned the median, and played again with the same group in a partner protocol. Payoffs were the amounts on the table, in pennies.
- 40 At the high equilibrium, the payoffs are larger and so the difference between the received payoff and δ times the foregone payoff will be larger than at the low equilibrium. (Numerically, a player who chooses 3 when the median is 3 earns \$0.60 and has a foregone payoff from 2 or 4 of $\$0.58 \cdot \delta$. The corresponding figures for a player choosing 12 are \$1.12 and $\$1.10 \cdot \delta$. The differences in received and foregone payoffs around 12 and around 3 are the same when $\delta = 1$, but the difference around 12 grows larger as δ falls (for example, for the self-tuning EWA estimate $\hat{\delta} = 0.69$, the differences are \$0.20 and \$0.36 for 3 and 12). Cumulating payoffs rather than averaging them 'blows up' the difference and produces sharper convergence at the high equilibrium.
- 41 They also had a session with $R = 10$ but in this session one subject sat out each round so we dropped it to avoid making an *ad hoc* assumption about learning in this unusual design. Each subject played ten times (and played with a different R for five more rounds; we use only the first ten rounds).
- 42 As λ rises, the QRE equilibria move sharply from smearing probability throughout the price range (for low λ) to a sharp spike at the equilibrium (higher λ). No intermediate λ can explain the combination of initial dispersion and sharp convergence at the end so the best-fitting QRE model essentially makes the Nash prediction.
- 43 In beauty contests and co-ordination games, payoffs depend on the mean or median of fairly large groups (7–9 except in 3-person entry games), so switching one subject's choice to the recommendation would probably not change the mean or median and hence would not change future behaviour much. In other games, players are usually paired randomly so the boomerang effect again is muted. We are currently redoing the analysis to simply compare profits of players whose choices frequently matched the recommendation with those who rarely did. This controls for the boomerang effect and also for a Lucas critique effect in which adopting recommendations would change the behaviour of others and hence the model parameters used to derive the recommendations. A more interesting correction is to run experiments in which one or more computerized subjects in fact use a learning model to make choices, and compare their performance with that of actual subjects.
- 44 For example, in the continental divide game, *ex-post* optimal payoffs would have been 892 (pennies per player) if players knew exactly what the median would be, and subjects in fact earned 837. EWA and self-tuning EWA generate simulated profits of 879–882, which is only an improvement of 5 per cent over 837, but is 80 per cent of the maximum possible improvement from actual payoffs to clairvoyant payoffs.
- 45 Partow and Schotter (1993), Mookerjee and Sopher (1994), Cachon and Camerer (1996).
- 46 Sophistication may also have the potential to explain why players sometimes move in the *opposite* direction to that predicted by adaptive models (Rapoport, 1999), and why measured beliefs do not match up well with those predicted by adaptive belief learning models (Nyarko and Schotter, 2002).
- 47 To truncate the belief hierarchy, the sophisticated players believe that the other sophisticated players, like themselves, believe there are α' sophisticates.

- 48 The gap between apparent sophistication and perceived sophistication shows the empirical advantage of separating the two. Using likelihood ratio tests, we can clearly reject both the rational expectations restriction $\alpha = \alpha'$ and the pure overconfidence restriction $\alpha' = 0$, although the differences in log-likelihood are not large.
- 49 Borrower subjects do not play consecutive sequences, which removes their incentive to repay in the eighth period of one sequence so they can get more loans in the first period of the next sequence.
- 50 To economize on computing, we search only paths of future actions that always have default following repay because the reverse behaviour (repay following default) generates a lower return.
- 51 Formally, $\hat{P}_L^j(a, k, v|j_{v-1}) = \hat{P}_L^{Loan}(a, k, v - 1|j_{v-1}) \cdot P_L^j(a, k, v|(Loan, j_{v-1})) + \hat{P}_L^{NoLoan}(a, k, v - 1|j_{v-1}) \cdot P_L^j(a, k, v|(NoLoan, j_{v-1}))$.
- 52 This is called 'observational learning' (see Duffy and Feltovich, 1999; Armentier, 2004). Without this assumption, the model learns far more slowly than do the lenders, so it is clear that they are learning from observing others.
- 53 The idea is to create an 'interim' attraction for Round t , $B_L^j(a, k, t)$, based on the attraction $A_L^j(a, k, t - 1)$ and payoff from Round t , then incorporate experience in Round $t + 1$ from previous sequences, transforming $B_L^j(a, k, t)$ into a final attraction $A_L^j(a, k, t)$. See Camerer *et al.* (2002) for details.
- 54 We use four separate λ s for honest borrowers, lenders, normal adaptive borrowers and teaching borrowers, an initial attraction for lending $A(0)$, and the spillover parameter σ and teaching proportion α .
- 55 See Camerer and Weigelt (1988); Palfrey and Rosenthal (1988); McKelvey and Palfrey (1992).
- 56 In our current extension of the Camerer *et al.* (2002) paper on strategic teaching, we impose a value of θ that was measured independently in a separate experiment on one-shot trust games. It is around 0.20, much lower than the estimate of 0.91, and when this measured number is used instead of letting θ be a free parameter, AQRE degrades rather badly.
- 57 One direction we are pursuing is to find designs or tests that distinguish the teaching and equilibrium updating approaches. The sharpest test is to compare behaviour in games with types that are fixed across sequences with types that are 'refreshed' independently in each period within a sequence. The teaching approach predicts similar behaviour in these two designs, but type-updating approaches predict that reputation formation dissolves when types are refreshed.
- 58 A fourth enterprise fits utility functions that reflect social preferences for fairness or equality. This important area is not discussed in this chapter.
- 59 See Smith *et al.*, 1988; Camerer and Weigelt, 1993; and Lei *et al.*, 2001.
- 60 See Sonsino *et al.*, 2000; Sovik, 2000.
- 61 The studies, in the order in which τ estimates are reported below, are Malcolm and Lieberman (1965), O'Neill (1987), Rapoport and Boebel (1992), Bloomfield (1994), Ochs (1995, games with 9 and 4 payoffs), Mookerjee and Sopher (1997), Rapoport and Almadoss (2000, $r = 8, 20$), Binmore *et al.* (2001), Tang (2001, games 3 and 1), games 1, 3, 4. Readers please let us know of published studies we have overlooked; we plan to include them in a later draft.
- 62 In each game, data from those $n - 1$ out of the n possible strategies with the most extreme predicted equilibrium probabilities are used to fit the models. Excluding the n th strategy reduces the dependence among data points, since all n frequencies (and predictions) obviously add up to one.

References

- Anderson, C. and Camerer, C. F. (2001) 'Experience-weighted Attraction Learning in Sender – Receiver Signaling Games', *Economic Theory*, 16, 3, 689–718.
- Andreoni, J. and Miller, J. (1993) 'Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence', *Economic Journal*, 103, 570–85.
- Armantier, Olivier (2004) 'Does Observation Influence Learning?', *Games and Economic Behavior*, 46, February, 221–39.
- Arthur, B. (1991) 'Designing Economic Agents That Act Like Human Agents: A Behavioral Approach to Bounded Rationality', *American Economic Review*, 81, 2, 353–9.
- Babcock, Linda and Loewenstein, George (1997) 'Explaining Bargaining Impasses: The Role of Self-serving Biases', *Journal of Economic Perspectives*, 11, 109–26.
- Babcock, L., Loewenstein, G., Issacharoff, S. and Camerer, C. F. (1995) 'Biased Judgments of Fairness in Bargaining', *American Economic Review*, 85, 1337–43.
- Bajari, P. and Hortacsu, A. (2003) 'Winner's Curse, Reserve Prices, and Endogeneous Entry: Empirical Insights from eBay', *RAND Journal of Economics*, Summer, 329–55.
- Binmore, K., Swierzbinski, J. and Proulx, C. (2001) 'Does Maximin Work? An Experimental Study', *Economic Journal*, 111, 445–64.
- Bloomfield, R. (1994) 'Learning a Mixed Strategy Equilibrium in the Laboratory', *Journal of Economic Behavior and Organization*, 25, 411–36.
- Blume, A., DeJong, D., Neumann, G. and Savin, N. (1999) 'Learning in Sender–Receiver Games', Working paper, University of Iowa.
- Brown, G. (1951) 'Iterative Solution of Games by Fictitious Play', in *Activity Analysis of Production and Allocation* T. C. Uoopmans (ed.), New York, John Wiley.
- Bush, R. and Mosteller, F. (1955) *Stochastic Models for Learning*, New York, John Wiley.
- Cabrales, A. and Garcia-Fontes, W. (2000) 'Estimating Learning Models with Experimental Data', Working paper, University of Pompeu Febr, Barcelona 501.
- Cachon, G. P. and Camerer, C. F. (1996) 'Loss-avoidance and Forward Induction in Experimental Coordination Games', *The Quarterly Journal of Economics*, 111, 165–94.
- Camerer, C. F. (1987) 'Do Biases in Probability Judgment Matter in Markets? Experimental Evidence', *American Economic Review*, 77, 981–97.
- Camerer, C. F. (1997) 'Progress in Behavioral Game Theory', *Journal of Economic Perspectives*, 11, 167–88.
- Camerer, C. F. (2003) *Behavioral Game Theory: Experiments on Strategic Interaction*, Princeton, NJ: Princeton University Press.
- Camerer, C. F. and Ho, T.-H. (1998) 'EWA Learning in Normal-form Games: Probability Rules, Heterogeneity and Time Variation', *Journal of Mathematical Psychology*, 42, 305–26.
- Camerer, C. F. and Ho, T.-H. (1999) 'Experience-weighted Attraction Learning in Normal-form Games', *Econometrica*, 67, 827–74.
- Camerer, C. F. and Ho, T.-H. (2001) 'Strategic Learning and Teaching in Games', in S. Hoch and H. Kunreuther (eds), *Wharton on Decision Making*, New York: Wiley.
- Camerer, C. F. and Lovallo, D. (1999) 'Overconfidence and Excess Entry: An Experimental Approach', *American Economic Review*, 89, 306–18.
- Camerer, C. F. and Weigelt, K. (1988) 'Experimental Tests of a Sequential Equilibrium Reputation Model', *Econometrica*, 56, 1–36.
- Camerer, C. F. and Weigelt, K. (1993) 'Convergence in Experimental Double Auctions for Stochastically Lived Assets', in D. Friedman and J. Rust (eds), *The Double Auction Market: Theories, Institutions and Experimental Evaluations*, Redwood City, Calif.: Addison-Wesley, 355–96.

- Camerer, C. F., Ho, T.-H. and Chong, J. K. (2002a) 'Sophisticated EWA Learning and Strategic Teaching in Repeated Games', *Journal of Economic Theory*, 104, 137–88.
- Camerer, C. F., Ho, T.-H. and Chong, J. K. (2002b) 'A Cognitive Hierarchy Theory of One-shot Games: Some Preliminary Results'. See <http://www.hss.caltech.edu/camerer/thinking2002.pdf>, accessed December.
- Camerer, C. F., Ho, T.-H. and Chong, J. K. (2003) 'Models of Thinking, Learning and Teaching in Games', *American Economic Review*, 93, May, 192–5.
- Camerer, C. F., Hsia, D. and Ho, T.-H. (2002) 'EWA Learning in Bilateral Call Markets', in A. Rapoport and R. Zwick (eds), *Experimental Business Research*, Dordrecht: Kluwer, 255–85.
- Camerer, C. F., Johnson, E., Sen, S. and Rymon, T. (1993) 'Cognition and Framing in Sequential Bargaining for Gains and Losses', in K. Binmore, A. Kirman and P. Tani (eds), *Frontiers of Game Theory*, Cambridge, Mass., MIT Press, 27–48.
- Capra, M. (1999) 'Noisy Expectation Formation in One-shot Games', Unpublished dissertation, University of Virginia.
- Capra, M., Goeree, J., Gomez, R. and Holt, C. (1999) 'Anomalous Behavior in a Traveler's Dilemma', *American Economic Review*, 89, 678–90.
- Cheung, Y.-W. and Friedman, D. (1997) 'Individual Learning in Normal Form Games: Some Laboratory Results', *Games and Economic Behavior*, 19, 46–76.
- Christiansen, M. H. and Chater, N. (1999) 'Toward a Connectionist Model of Recursion in Human Linguistic Performance', *Cognitive Science*, 23, 157–205.
- Clark, K. and Sefton, M. (1999) 'Matching Protocols in Experimental Games', Working paper, University of Manchester. See <http://nt2.ec.man.ac.uk/ses/discpap/downloads/9917.pdf>.
- Cooper, D. and Kagel, John (2003a) 'Lessons Learned: Generalizing Learning Across Games', *American Economic Review*, 93, May, 202–7.
- Cooper, D. and Kagel, John (2003b) 'Are Two Heads Better Than One? Team versus Individual Play in Signaling Games', Working paper, Case Western, July. See <http://www.econ.ohio-state.edu/kagel/teamsD71.1.pdf>.
- Cooper, D. and Van Huyck, J. (2003) 'Evidence on the Equivalence of the Strategic and Extensive Form Representation of Games', *Journal of Economic Theory*, 110, 2, June, 290–308.
- Costa-Gomes, Miguel and Weizsäcker, G. (2003) 'Stated Beliefs and Play in Normal-form Games', Working paper, ISER and Harvard, November. See <http://kuznets.fas.harvard.edu/weizsack/papers/CGW3.pdf>.
- Costa-Gomes, M., Crawford, V. and Broseta, B. (2001) 'Cognition and Behavior in Normal-form Games: An Experimental Study', *Econometrica*, 69, 1193–235.
- Crawford, V. (1997) 'Theory and Experiment in the Analysis of Strategic Interactions', in D. Kreps and K. Wallis (eds), *Advances in Economics and Econometrics: Theory and Applications*, Seventh World Congress, Vol. 1, Cambridge University Press.
- Crick, F. (1988) *What Mad Pursuit?*, New York: Sloan Foundation.
- Croson, R. T. A. (2000) 'Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play', *Journal of Economic Behavior and Organization*, 41, 3, 299–314.
- Cross, J. (1983) *A Theory of Adaptive Learning Economic Behavior*, New York, Cambridge University Press.
- Duffy, J. and Feltovich, N. (1999) 'Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study', *International Journal of Game Theory*, 28, 131–52.
- Erev, I. and Roth, A. (1998) 'Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed-strategy Equilibria', *The American Economic Review*, 88, 848–81.

- Erev, I., Bereby-Meyer, Y. and Roth, A. (1999) 'The Effect of Adding a Constant to All Payoffs: Experimental Investigation, and a Reinforcement Learning Model with Self-Adjusting Speed of Learning', *Journal of Economic Behavior and Organization*, 39, 111–28.
- Fudenberg, D. and Levine, D. (1989) 'Reputation and Equilibrium Selection in Games with a Patient Player', *Econometrica*, 57, 759–78.
- Fudenberg, D. and Levine, D. (1998) *The Theory of Learning in Games*, Boston, Mass., MIT Press.
- Goeree, J. K. and Holt, C. A. (1999) 'Stochastic Game Theory: For Playing Games, Not Just for Doing Theory', *Proceedings of the National Academy of Sciences*, 96, 10564–7.
- Goeree, J. K. and Holt, C. A. (2001) 'Ten Little Treasures of Game Theory, and Ten Intuitive Contradictions', *American Economic Review*, 91, December, 1402–22.
- Goeree, J. K. and Holt, C. A. (2004) 'A Theory of Noisy Introspection', *Games and Economic Behavior*, 46, February, 365–82.
- Güth, Werner (2000) 'Boundedly Rational Decision Emergence – A General Perspective and Some Selective Illustrations', *Journal of Economic Psychology*, 21, 433–58.
- Harley, C. (1981) 'Learning the Evolutionary Stable Strategies', *Journal of Theoretical Biology*, 89, 611–33.
- Heller, D. and Sarin, R. (2000) 'Parametric Adaptive Learning', Working paper, University of Chicago.
- Ho, T.-H. and Chong, J. K. (2003) 'A Parsimonious Model of SKU Choice', *Journal of Marketing Research*, 40, 351–65.
- Ho, T.-H. and Weigelt, K. (1996) 'Task Complexity, Equilibrium Selection, and Learning: An Experimental Study', *Management Science*, 42, 659–79.
- Ho, T.-H., Camerer, C. F. and Chong, J. K. (2001) 'Economic Value of EWA Lite: A Functional Theory of Learning in Games', Working paper, University of Pennsylvania. See <http://www.hss.caltech.edu/camerer/camerer.html>.
- Ho, T.-H., Camerer, C. F. and Weigelt, K. (1998) 'Iterated Dominance and Iterated Best Response in Experimental 'p-Beauty Contests'', *American Economic Review*, 88, 947–69.
- Ho, T.-H., Wang, X. and Camerer, C. F. (1999) 'Individual Differences and Payoff Learning in Games', Working paper, University of Pennsylvania.
- Holt, Debra (1999) 'An Empirical Model of Strategic Choice with an Application to Coordination', *Games and Economic Behavior*, 27, 86–105.
- Hopkins, E. (2002) 'Two Competing Models of How People Learn in Games', *Econometrica*, 70, 2141–66.
- Huberman, G. and Rubinstein, A. (2000) 'Correct Belief, Wrong Action and a Puzzling Gender Difference', Working paper, Tel Aviv University.
- Jehiel, P. (2002) 'Predicting by Analogy and Limited Foresight in Games', Working paper, University College London.
- Jehiel, P. (forthcoming) Bounded Rationality and Imperfect Learning: Game Theory versus AI', *Greek Economic Review*.
- Johnson, E. J., Camerer, C. F., Sen, S. and Rymon, T. (2002) 'Detecting Backward Induction in Sequential Bargaining Experiments', *Journal of Economic Theory*, 104, May, 16–47.
- Jung, Y. J., Kagel, J. H. and Levin, D. (1994) 'On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain-store Game', *RAND Journal of Economics*, 25, 72–93.
- Kahneman, D. (1988) 'Experimental Economics: A Psychological Perspective', in R. Tietz, W. Albers and R. Selten (eds), *Bounded Rational Behavior in Experimental Games and Markets*, New York, Springer-Verlag, 11–18.

- Kahneman, Daniel (2003) 'Maps of Bounded Rationality: Psychology for Behavioral Economics', *American Economic Review*, 93, December, 1449–75.
- Kaufman, H. and Becker, G. M. (1961) 'The Empirical Determination of Game-theoretical Strategies', *Journal of Experimental Psychology*, 61, 462–8.
- Kocher, M. G. and Sutter, M. (forthcoming) 'When the 'Decision Maker' Matters: Individual versus Team Behavior in Experimental "Beauty-contest" Games', *Economic Journal*.
- Kreps, D. and Wilson, R. (1982) 'Reputation and Imperfect Information', *Journal of Economic Theory*, 27, 253–79.
- Lei, V., Noussair, C. and Plott, C. (2001) 'Non-speculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality', *Econometrica*, 69, 813–59.
- Lucas, R. G. (1986) 'Adaptive Behavior and Economic Theory', *Journal of Business*, 59, October, S401–S426.
- Marcet, Albert and Nicolini, Juan P. (2003) 'Recurrent Hyperinflations and Learning', *American Economic Review*, 93, December, 1476–98.
- Mailath, G. (1998) 'Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory', *Journal of Economic Literature*, 36, 1347–74.
- Malcolm, D. and Lieberman, B. (1965) 'The Behavior of Responsive Individuals Playing a Two-person Zero-sum Game Requiring the Use of Mixed Strategies', *Psychonomic Science*, 2, 373–4.
- McAllister, P. H. (1991) 'Adaptive Approaches to Stochastic Programming', *Annals of Operations Research*, 30, 45–62.
- McKelvey, R. D. and Palfrey, T. R. (1992) 'An Experimental Study of the Centipede Game', *Econometrica*, 60, 803–36.
- McKelvey, R. D. and Palfrey, T. R. (1995) 'Quantal Response Equilibria for Normal-form Games', *Games and Economic Behavior*, 7, 6–38.
- McKelvey, R. D. and Palfrey, T. R. (1998) 'Quantal Response Equilibria for Extensive-form Games', *Experimental Economics*, 1, 9–41.
- Mookerjee, D. and Sopher, B. (1994) 'Learning Behavior in an Experimental Matching Pennies Game', *Games and Economic Behavior*, 7, 62–91.
- Mookerjee, D. and Sopher, B. (1997) 'Learning and Decision Costs in Experimental Constant-sum Games', *Games and Economic Behavior*, 19, 97–132.
- Nagel, R. (1995) 'Experimental Results on Interactive Competitive Guessing', *The American Economic Review*, 85, 1313–26.
- Nagel, R. (1999) 'A Review of Beauty Contest Games', in D. Budescu, I. Erev and R. Zwick (eds), *Games and Human Behavior: Essays in honor of Amnon Rapoport*, Lawrence Erlbaum Associates, Mahwah NJ, 105–42.
- Nagel, R., Bosch-Domenech, A., Satorra, A. and Garcia-Montalvo, J. (2002) 'One, Two, (Three), Infinity: Newspaper and Lab Beauty-contest Experiments', *American Economic Review*.
- Neral, J. and Ochs, J. (1992) 'The Sequential Equilibrium Theory of Reputation Building: A Further Test', *Econometrica*, 60, 1151–69.
- Nyarko, Y. and Schotter, A. (2002) 'An Experimental Study of Belief Learning Using Elicited Beliefs', *Econometrica*, 70, May, 971–1005.
- Ochs, J. (1995) 'Games with Unique, Mixed Strategy Equilibria: An Experimental Study', *Games and Economic Behavior*, 10, 202–17.
- Ochs, J. (1999) 'Entry in Experimental Market Games', in D. Budescu, I. Erev and R. Zwick (eds), *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, Mahwah, NJ, Lawrence Erlbaum Associates.

- O'Neill, B. (1987) 'Nonmetric Test of the Minimax Theory of Two-person Zero-sum Games', *Proceedings of the National Academy of Sciences*, 84, 2106–9.
- Palfrey, T. R. and Rosenthal, H. (1988) 'Private Incentives in Social Dilemmas: The Effects of Incomplete Information and Altruism', *Journal of Public Economics*, 35, 309–32.
- Partow, J. and Schotter, A. (1993) 'Does Game Theory Predict Well for the Wrong Reasons? An Experimental Investigation', Working paper No. 93–46, C. V. Starr Center for Applied Economics, New York University.
- Plott, C. R. (1986) 'Dimensions of Parallelism: Some Policy Applications of Experimental Methods', in A. E. Roth (ed.), *Laboratory Experimentation in Economics: Six Points of View*, Cambridge University Press.
- Plott, C. R. (1997) 'Laboratory Experimental Testbeds: Applications to the PCS Auctions', *Journal of Economics and Management Strategy*, 6, 605–38.
- Prendergast, C. (1999) 'The Provision of Incentives in Firms', *Journal of Economic Literature*, 37, March, 7–63.
- Rapoport, A. and Amaldoss, W. (2000) 'Mixed Strategies and Iterative Elimination of Strongly Dominated Strategies: An Experimental Investigation of States of Knowledge', *Journal of Economic Behavior and Organization*, 42, 483–521.
- Rapoport, A. and Boebel, R. B. (1992) 'Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis', *Games and Economic Behavior*, 4, 261–83.
- Rapoport, A. and Erev, I. (1998) 'Coordination, "Magic", and Reinforcement Learning in a Market Entry Game', *Games and Economic Behavior*, 23, 146–75.
- Rapoport, A., Lo, K.-C. and Zwick, R. (1999) 'Choice of Prizes Allocated by Multiple Lotteries with Endogenously Determined Probabilities', Working paper, University of Arizona, Department of Management and Policy.
- Rassenti, S. J., Smith, V. L. and Wilson, B. J. (2001) 'Turning Off the Lights', *Regulation*, Fall, 70–6.
- Roth, A. and Erev, I. (1995) 'Learning in Extensive-form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term', *Games and Economic Behavior*, 8, 164–212.
- Roth, A. and Xing, X. (1994) 'Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions', *American Economic Review*, 84, 992–1044.
- Roth, A., Barron, G., Erev, I. and Slonim, R. (2000) 'Equilibrium and Learning in Economic Environments: The Predictive Value of Approximations', Working paper, Harvard University.
- Salmon, T. (2001) 'An Evaluation of Econometric Models of Adaptive Learning', *Econometrica*, 69, 1597–628.
- Salmon, T. (2003) 'Evidence for "Learning to Learn" Behavior in Normal-form Games', Paper, Florida State University, August. See <http://garnet.acns.fsu.edu/tsalmon/EvidV2.pdf>.
- Sargent, T. (1999) *The Conquest of American Inflation*, Princeton, NJ, Princeton University Press.
- Schelling, T. (1960) *The Strategy of Conflict*, Cambridge Mass Harvard University Press.
- Schotter, A., Weigelt, K. and Wilson, C. (1994) 'A Laboratory Investigation of Multiperson Rationality and Presentation Effects', *Games and Economic Behavior*, 6, 445–68.
- Seale, D. A. and Rapoport, A. (2000) 'Elicitation of Strategy Profiles in Large Group Coordination Games', *Experimental Economics*, 3, 153–79.
- Selten, R. (1978) 'The Chain Store Paradox', *Theory and Decision*, 9, 127–59.
- Reinhard, Selten, (1998) 'Features of Experimentally Observed Bounded Rationality', *European Economic Review*, 42, 413–36.

- Selten, R. and Stoecker, R. (1986) 'End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach', *Journal of Economic Behavior and Organization*, 7, 47–70.
- Smith, V. L., Suchanek, G. and Williams, A. (1988) 'Bubbles, Crashes and Endogeneous Expectations in Experimental Spot Asset Markets', *Econometrica*, 56, 1119–51.
- Sonsino, D., Erev, I. and Gilat, S. (2000) 'On the Likelihood of Repeated Zero-sum Betting by Adaptive (Human) Agents', Technion, Israel Institute of Technology.
- Sovik, Y. (2000) 'Impossible Bets: An Experimental Study', Mimeo University of Oslo Department of Economics.
- Stahl, D. O. (1996) 'Boundedly Rational Rule Learning in a Guessing Game', *Games and Economic Behavior*, 16, 303–30.
- Stahl, Dale O. (1999) 'Sophisticated Learning and Learning Sophistication', Working paper, University of Texas at Austin.
- Stahl, D. O. (2000) 'Local Rule Learning in Symmetric Normal-form Games: Theory and Evidence', *Games and Economic Behavior*, 32, 105–38.
- Stahl, D. O. and Wilson, P. (1995) 'On Players Models of Other Players: Theory and Experimental Evidence', *Games and Economic Behavior*, 10, 213–54.
- Sundali, J. A., Rapoport, A. and Seale, D. A. (1995) 'Coordination in Market Entry Games with Symmetric Players', *Organizational Behavior and Human Decision Processes*, 64, 203–18.
- Sutton, R. and Barto, A. (1998) *Reinforcement Learning: An Introduction*, Boston, Mass., MIT Press.
- Svensson, L. E. O. (2001) 'Comments on Michael Woodford Paper', presented at Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps', Columbia University, 5–6 October.
- Tang, F.-F. (2001) 'Anticipatory Learning in Two-person Games: Some Experimental Results', *Journal of Economic Behavior and Organization*, 44, 221–32.
- Timmerman, A. G. (1993) 'How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices', *Quarterly Journal of Economics*, 108, November, 1135–45.
- J. Tirole (1985) 'Asset Bubbles and Overlapping Generations', *Econometrica*, 53, 1071–100 (reprinted 1499–528).
- Van Damme, E. (1999) 'Game Theory: The Next Stage', in L. A. Gerard-Varet, Alan P. Kirman and M. Ruggiero (eds), *Economics Beyond the Millennium*, Oxford University Press, 184–214.
- Van Huyck, J., Battalio, R. and Beil, R. (1990) 'Tacit Cooperation Games, Strategic Uncertainty, and Coordination Failure', *The American Economic Review*, 80, 234–48.
- Van Huyck, J., Battalio, R. and Rankin, F. W. (2001) 'Selection Dynamics and Adaptive Behavior Without Much Information', Working paper, Texas A & M Department of Economics.
- Van Huyck, J., Cook, J. and Battalio, R. (1997) 'Adaptive Behavior and Coordination Failure', *Journal of Economic Behavior and Organization*, 32, 483–503.
- Von Neumann, J. and Morgenstern, O. (1944) *The Theory of Games and Economic Behavior*, Princeton, NJ, Princeton University Press.
- Warglien, M., Devetag, M. G. and Legrenzi, P. (1998) 'Mental Models and Naive Play in Normal Form Games', Working paper, Universita Ca' Foscari di Venezia.
- Watson, J. (1993) 'A "Reputation" Refinement without Equilibrium', *Econometrica*, 61, 199–205.
- Watson, J. and Battigali, P. (1997) 'On "Reputation" Refinements with Heterogeneous Beliefs', *Econometrica*, 65, 363–74.

- Weber, Roberto (forthcoming) '“Learning” with No Feedback in a Competitive Guessing Game', *Games and Economic Behavior*.
- Weibull, J. (1995) *Evolutionary Game Theory*, Cambridge, Mass., MIT Press.
- Weizsäcker, G. (2003) 'Ignoring the Rationality of Others: Evidence from Experimental Normal-form Games', *Games and Economic Behavior*, 44, 145–71.
- Wilcox, Nat (2003) 'Heterogeneity and Learning Models', Working paper, University of Houston. See http://www.uh.edu/~wilcox/papers/hetero_and_learning_W.pdf.
- Woodford, M. (2001) 'Imperfect Common Knowledge and the Effects of Monetary Policy', Paper presented at 'Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps', Columbia University, 5–6 October.