

CHAPTER 10

Behavioral Game Theory Experiments and Modeling

Colin F. Camerer^{*}, Teck-Hua Ho[†]

^{*}California Institute of Technology, Pasadena, CA, USA

[†]University of California, Berkeley, Berkeley, CA, USA

Contents

10.1. Introduction	518
10.2. Cognitive Hierarchy and Level-k Models	520
10.2.1 <i>P</i> -beauty contest	523
10.2.2 Market entry games	526
10.2.3 LUPI lottery	528
10.2.4 Summary	530
10.3. Quantal Response Equilibrium	530
10.3.1 Asymmetric hide-and-seek game	531
10.3.2 Maximum value auction	533
10.4. Learning	535
10.4.1 Parametric EWA learning: interpretation, uses, and limits	536
10.4.2 fEWA functions	539
10.4.2.1 <i>The change-detector function</i> $\phi_i(t)$	540
10.4.2.2 <i>The attention function</i> , $\delta_{ij}(t)$	542
10.4.3 fEWA predictions	545
10.4.4 Example: mixed strategy games	548
10.4.5 Summary	551
10.5. Sophistication and Teaching	551
10.5.1 Sophistication	552
10.5.2 Strategic teaching	554
10.5.3 Summary	559
10.6. Sociality	560
10.6.1 Public goods	560
10.6.2 Public goods with punishment	561
10.6.3 Negative reciprocity: ultimatums	562
10.6.4 Impure altruism and social image: dictator games	564
10.6.5 Summary	566
10.7. Conclusion	566
References	567

Abstract

This chapter reviews recent experimental data testing game theory and behavioral models that have been inspired to explain those data. The models fall into four groups: in cognitive hierarchy or level- k models, the assumption of equilibrium is relaxed by assuming agents have beliefs about other agents who do less reasoning (i.e., some are non-strategic, and others are more strategic and understand they are playing non-strategic players). A different approach, quantal response equilibrium, retains equilibrium expectations but adds stochastic response (of which players are aware). Learning theories explain choices in games played repeatedly as a result of past actions and payoffs that were received (as in classical reinforcement) or foregone payoffs (model-directed learning). Finally, many studies reject the joint hypothesis of equilibrium expectations and optimization, along with self-interest in valuing outcomes. Social preference models have emerged to explain these data, capturing concepts like inequity-aversion, reciprocity, and social image.

Keywords: Behavioral game theory, Cognitive hierarchy, Level- k , Learning, Nonequilibrium, Bounded rationality

JEL Codes: C7, C9, D7, D8

10.1. INTRODUCTION

For explaining individual decisions, rationality—in the sense of accurate belief and optimization—may still be an adequate approximation even if a modest percentage of players violate the theory. But game theory is different. Players' fates are intertwined. The presence of players who do not have accurate belief or optimize can dramatically change what rational players should do. As a result, what a population of players is likely to do when some do not have accurate belief or optimize can only be predicted by an analysis which explicitly accounts for bounded rationality as well, preferably in a precise way. This chapter is about what has been learned about boundedly rational strategic behavior from hundreds of experimental studies (and some field data).

In the experiments, equilibrium game theory is almost always the benchmark model being tested. However, the frontier has moved well beyond simply comparing actual behaviors and equilibrium predictions, because that comparison has inspired several types of behavioral models. Therefore, the chapter is organized around precise behavioral models of boundedly rational choice, learning and strategic teaching, and social preferences. We focus selectively on several example games which are of economic interest and explain how these models work.

We focus on five types of models

1. Cognitive hierarchy (CH) model which captures players' beliefs about steps of thinking, using one parameter to describe the average step of strategic thinking (level- k (LK) models are closely related). These models are designed to predict one-shot games or initial conditions in a repeated game.

2. A noisy optimization model called quantal response equilibrium (QRE). Under QRE, players are allowed to make small mistakes but they always have accurate beliefs about what other players will do.
3. A learning model (called experience-weighted attraction learning (EWA)) to compute the path of equilibration. The EWA learning algorithm generalizes both fictitious play and reinforcement models. EWA can predict the time path of play as a function of the initial conditions (perhaps supplied by CH model).
4. An extension of the learning models to include sophistication (understanding how others learn) as well as strategic teaching, and nonequilibrium reputation-building. This class of models allows the population to have sophisticated players who actively influence adaptive players' learning paths to benefit themselves.
5. Models of how social preferences map monetary payoffs (controlled in an experiment) into utilities and behavior.

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

Precision: Because game theory predictions are sharp, it is not hard to spot likely deviations and counterexamples. 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 distill the deviations into an alternative theory that is similarly precise as standard theory and can be widely applied. We favor specifications that use one or two free parameters to express crucial elements of behavioral flexibility. This approach often embeds standard equilibrium as a parametric special case of general theory. It also allows sensible measures of individual and group differences, through parameter variation. We also prefer to let data, rather than intuition, specify parameter values.

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 enables knowledge to cumulate worldwide. Behavioral models of games are also meant to be general, too, in the sense that the same models can be applied to many games with minimal customization. Keep in mind that this desire for universal application is not held in all social sciences. For example, many researchers in psychology believe that behavior is so context-specific that it is might be too challenging to create a common theory that applies to all contexts. Our view is that we can't know whether general theories are hopeless until we try to apply them broadly.

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 Nobel Prize-winning physicist Murray Gell-Mann said, "Imagine how difficult physics

would be if electrons could think.” It is even harder if we don’t watch what ‘electrons’ do when interacting.

Our insistence on empirical discipline is shared by others. Van Damme (1999) thought the same:

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.

Experiments play a privileged role in testing game theory. Game-theoretic predictions are notoriously sensitive to what players know, when they move, and how they value outcomes. Laboratory environments provide crucial control of all these variables (see Camerer, 2003; Crawford, 1997). As in other lab sciences, the idea is to use lab control to sort out which theories work well and which don’t, then later use them to help understand patterns in naturally occurring data. In this respect, behavioral game theory resembles data-driven fields like labor economics or finance more than equilibrium game theory.

Many behavioral game theory models also circumvent two long-standing problems in equilibrium game theory: refinement and selection. These models “automatically” refine sets of Bayesian-Nash equilibria (BNE) because they allow all events to occur with positive probability, and hence Bayes’ rule can be used in all information sets. Some models (e.g., CH and LK) also automatically avoid multiplicity of equilibria by making a single statistical prediction. Surprisingly, assuming *less* rationality on players therefore can lead to *more* precision (as noted previously by Lucas, 1986).

10.2. COGNITIVE HIERARCHY AND LEVEL-K MODELS

CH and LK models are designed to predict behavior in one-shot games and to provide initial conditions for models of learning.

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:

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{j'=1}^{m_i} e^{\lambda \cdot A_i^{j'}(t)}} \quad [10.1]$$

where λ is the response sensitivity.¹ Since CH and LK models are designed to predict strategic behaviors in only one-shot games, we focus mainly on $P_i^j(1)$ (i.e., no learning). We shall use the same framework for learning models too and the model predictions there will go beyond one period.

We model thinking by characterizing the number of steps of iterated thinking that subjects do, their beliefs, and their decision rules.² Players using zero steps of thinking, do not reason strategically at all. We think these players are using simple low-effort heuristics, such as choosing salient strategies (cf. [Shah and Oppenheimer, 2008](#)). In games which are physically displayed (e.g., battlefields) salience might be based on visual perception (e.g., [Itti and Koch, 2000](#)). In games with private information, a strategy choice that matches an information state might be salient (e.g., bidding one's value or signal in an auction). Randomizing among all strategies is also a reasonable step-0 heuristic when no strategy is particularly salient.

Players who do one step of thinking believe they are playing against step 0 types. Proceeding inductively, players who use k steps think all others use from zero to $k - 1$ steps. Since they do not imagine same- and higher-step types there is missing probability; we assume beliefs are normalized to adjust for this missing probability. Denote beliefs of level- k players about the proportion of step h players by $g_k(h)$. In CH, $g_k(h) = f(h) / \sum_{k'=0}^{k-1} f(k')$ for $h \leq k - 1$ and $g_k(h) = 0$ for $h \geq k$. In LK, $g_k(k - 1) = 1$. The LK model is easier to work with analytically for complex games.

It is useful to ask why the number of steps of thinking might be limited. One possible answer comes from psychology. Steps of thinking strain “working memory,” because higher step players need to keep in mind more and more about what they believe lower steps will do as they form beliefs. Evidence that cognitive ability is correlated with more thinking steps is consistent with this explanation ([Camerer et al., 2004](#); [Devetag and Warglien, 2003](#); [Gill and Prowse, 2012](#)).

However, it is important to note that making a step k choice does not logically imply a limit on level of thinking. In CH, players are essentially defined by their beliefs about others, not by their own cognitive capabilities. For example, a step 2 player believes others use 0 or 1 steps. It is possible that such players are capable of even higher-step reasoning (e.g., choosing Nash equilibria) if they thought their opponents were of higher steps. CH and level- k models are not meant to permit separation of beliefs and “reasoning skill,” per se (though see [Kneeland, 2013](#) for evidence).

An appealing intuition for why strategic thinking is limited is that players endogenously choose whether to think harder. The logical challenge in such an approach is this: A player who fully contemplated the benefit from doing one more thinking

¹ Note the timing convention—attractions are defined *before* a period of play; so the initial attractions $A_i^j(0)$ determine choices in period 1, and so forth.

² This concept was first studied by [Stahl and Wilson \(1995\)](#) and [Nagel \(1995\)](#), and later by [Ho et al. \(1998\)](#).

step would have to derive the optimal strategy after the additional thinking. [Alaoui and Penta \(2013\)](#) derive an elegant axiomatization of such a process by assuming that players compute the benefit from an extra step of thinking based on its maximum payoff. They also calibrate their approach to data of [Goeree and Holt \(2001\)](#) with some success.

Of course, it is also possible to allow players to change their step k choice. Indeed, [Ho and Su \(2013\)](#) and [Ho et al. \(2013\)](#) make CH and level- k models dynamic by allowing players to update their beliefs of other players' steps of thinking in sequential games. In their dynamic level- k model, players not only choose rules based on their initial guesses of others' steps (like CH and level- k) but also use historical plays to improve their guesses. Their dynamic level- k model captures two systematic violations of backward induction in centipede games, limited induction (i.e., people violate backward induction more in sequential games with longer decision trees) and repetition unraveling (i.e., choices unravel towards backward induction outcomes over time) and explains learning in p-beauty contests and sequential bargaining games.

The key challenge in thinking steps models is pinning down the frequencies of players using different numbers of thinking steps. A popular nonparametric approach is to let $f(k)$ be free parameters for a limited number of steps k ; often the three steps 0, 1, and 2 will fit adequately. Instead, we assume those frequencies have a Poisson distribution with mean and standard deviation τ (the frequency of step k types is $f(k) = \frac{e^{-\tau} \tau^k}{k!}$). Then τ is an index of aggregate bounded rationality.

The Poisson distribution has only one free parameter (τ); it generates “spikes” in predicted distributions reflecting individual heterogeneity (some other approaches do not reflect heterogeneity³); and for typical τ values the Poisson frequency of step types is roughly similar to estimates of nonparametric frequencies (see [Ho et al., 1998](#); [Nagel et al., 1999](#); [Stahl and Wilson, 1995](#)).

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

$$A_i^j(0|k) = \sum_{h=1}^{m-i} \pi_i(s_i^j, s_{-i}^h) \cdot \left\{ \sum_{c=0}^{k-1} \left[\frac{f(c)}{\sum_{k'=0}^{k-1} f(c)} \cdot P_{-i}^h(1|k') \right] \right\} \quad [10.2]$$

where $A_i^j(0|k)$ and $P_i^j(1|k')$ are the attraction of step k in period 0 and the predicted choice probability of lower step k' in period 1. Hence, the predicted probability of level K in period 1 is given by:

$$P_i^j(1|K) = \frac{e^{\lambda \cdot A_i^j(0|K)}}{\sum_{j'=1}^{m_i} e^{\lambda \cdot A_i^{j'}(0|K)}}. \quad [10.3]$$

³ Alternatives include the theory of noisy expectation by [Capra \(1999\)](#) and the closely related theory of “noisy introspection” by [Goeree and Holt \(1999\)](#). Both models do not accommodate heterogeneity.

Next we apply CH to three types of games: dominance-solvable p -beauty contests, market entry, and LUPI lottery choices. Note that CH and LK have also been applied to literally hundreds of other games of many structures (including private information) (see Crawford et al., 2013).

10.2.1 P -beauty contest

In a famous passage, Keynes (1936) likens the stock market to a newspaper contest in which people guess which faces others will judge to be the most beautiful. He writes:

It is not the case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree, where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth, and higher degrees [p. 156].

The essence of Keynes's observation is captured in a game in which players are asked to pick numbers from 0 to 100, and the player whose number is closest to p times the average number wins a prize. Suppose $p = 2/3$ (a value used in many experiments). Equilibrium theory predicts each contestant will reason as follows: "Even if all the other players guess 100, I should guess no more than $2/3$ times 100, or 67. Assuming that the other contestants reason similarly, however, I should guess no more than 45. . ." and so on, finally concluding that the only rational and consistent choice for all the players is zero. When the beauty contest game is played in experimental settings, the group average is typically between 20 and 35. Apparently, some players are not able to reason their way to the equilibrium value, or they assume that others are unlikely to do so. If the game is played multiple times with the same group, the average moves close to 0.

Table 10.1 shows data from 24 p -beauty contest games, with various p and conditions. Estimates of τ for each game were chosen to minimize the (absolute) difference between the predicted and actual mean of chosen numbers. The table is ordered from top to bottom by the mean number chosen. The first seven lines show games in which the equilibrium is not zero; in all the others the equilibrium is zero.⁴

The first four columns describe the game or subject pool, the source, group size, and total sample size. The fifth and sixth columns show the Nash equilibrium and the difference between the equilibrium and the average choice. The middle three columns show the mean, standard deviation, and the mode in the data. The mean choices are generally far off from the equilibrium; they choose numbers that are too low when the equilibrium is high (first six rows) and the numbers that are too high when the equilibrium is low (lower rows). The rightmost six columns show the estimate of τ from the Poisson-CH model, and the mean, prediction error, standard deviation, and mode predicted by

⁴ In games in which the equilibrium is not zero, players are asked to choose a number that is closest to p times the average number + a nonzero constant.

Table 10.1 Data and ch estimates of τ in various p-beauty contest games.

Subject pool or game	Source ¹	Group		Nash equil'm	Pred'n error	Data			Fit of CH model			Bootstrapped		
		size	size			Mean	Std Dev	Mode	τ	Mean	Error	Std Dev	Mode	90% C.I.
$p = 1.1$	HCW (98)	7	69	200	47.9	152.1	23.7	150	0.10	151.6	-0.5	28.0	165	(0.0,0.5)
$p = 1.3$	HCW (98)	7	71	200	50.0	150.0	25.9	150	0.00	150.4	0.5	29.4	195	(0.0,0.1)
High \$	CHW	7	14	72	11.0	61.0	8.4	55	4.90	59.4	-1.6	3.8	61	(3.4,4.9)
Male	CHW	7	17	72	14.4	57.6	9.7	54	3.70	57.6	0.1	5.5	58	(1.0,4.3)
Female	CHW	7	46	72	16.3	55.7	12.1	56	2.40	55.7	0.0	9.3	58	(1.6,4.9)
Low \$	CHW	7	49	72	17.2	54.8	11.9	54	2.00	54.7	-0.1	11.1	56	(0.7,3.8)
$0.7(M + 18)$	Nagel (98)	7	34	42	-7.5	49.5	7.7	48	0.20	49.4	-0.1	26.4	48	(0.0,1.0)
PCC	CHC (new)	2	24	0	-54.2	54.2	29.2	50	0.00	49.5	-4.7	29.5	0	(0.0,0.1)
$p = 0.9$	HCW (98)	7	67	0	-49.4	49.4	24.3	50	0.10	49.5	0.0	27.7	45	(0.1,1.5)
PCC	CHC (new)	3	24	0	-47.5	47.5	29.0	50	0.10	47.5	0.0	28.6	26	(0.1,0.8)
Caltech board	Camerer	73	73	0	-42.6	42.6	23.4	33	0.50	43.1	0.4	24.3	34	(0.1,0.9)
$p = 0.7$	HCW (98)	7	69	0	-38.9	38.9	24.7	35	1.00	38.8	-0.2	19.8	35	(0.5,1.6)
CEOs	Camerer	20	20	0	-37.9	37.9	18.8	33	1.00	37.7	-0.1	20.2	34	(0.3,1.8)
German students	Nagel (95)	14-16	66	0	-37.2	37.2	20.0	25	1.10	36.9	-0.2	19.4	34	(0.7,1.5)
70 yr olds	Kovalchik	33	33	0	-37.0	37.0	17.5	27	1.10	36.9	-0.1	19.4	34	(0.6,1.7)
US high school	Camerer	20-32	52	0	-32.5	32.5	18.6	33	1.60	32.7	0.2	16.3	34	(1.1,2.2)
Econ PhDs	Camerer	16	16	0	-27.4	27.4	18.7	N/A	2.30	27.5	0.0	13.1	21	(1.4,3.5)
1/2 mean	Nagel (98)	15-17	48	0	-26.7	26.7	19.9	25	1.50	26.5	-0.2	19.1	25	(1.1,1.9)
Portfolio mgrs	Camerer	26	26	0	-24.3	24.3	16.1	22	2.80	24.4	0.1	11.4	26	(2.0,3.7)
Caltech students	Camerer	17-25	42	0	-23.0	23.0	11.1	35	3.00	23.0	0.1	10.6	24	(2.7,3.8)
Newspaper	Nagel (98)	3696, 1460, 2728	7884	0	-23.0	23.0	20.2	1	3.00	23.0	0.0	10.6	24	(3.0,3.1)
Caltech	CHC (new)	2	24	0	-21.7	21.7	29.9	0	0.80	22.2	0.6	31.6	0	(4.0,1.4)
Caltech	CHC (new)	3	24	0	-21.5	21.5	25.7	0	1.80	21.5	0.1	18.6	26	(1.1,3.1)
Game theorists	Nagel (98)	27-54	136	0	-19.1	19.1	21.8	0	3.70	19.1	0.0	9.2	16	(2.8,4.7)
								mean	1.30					
								median	1.61					

Note 1: HCW (98) is Ho et al. (1998); CHC are new data from Camerer, Ho, and Chong; CHW is Camerer et al. (unpublished); Kovalchik is unpublished data collected by Stephanie Kovalchik

the best-fitting estimate of τ , and the 90% confidence interval of τ estimated from a randomized resampling (bootstrap) procedure.

There are several interesting patterns in Table 10.1. The prediction errors of the mean (column 13, “error”) are extremely small, less than .6 in all but two cases. This is no surprise since τ is estimated (separately in each row) to minimize this prediction error. The pleasant surprise is that the predicted standard deviations and the modal number choices which result from the error-minimizing estimate of τ are also fairly close (across rows, the correlation of the predicted and actual standard deviation is 0.72) even though τ ’s were not chosen to match these moments. The values of τ have a median and mean across rows of 1.30 and 1.61. The confidence intervals have a range of about one in samples of reasonable size (50 subjects or more).

Note that nothing in the Poisson-CH model, per se, requires τ to be fixed across games or subject pools, or across details of how games are presented or choices are elicited. Outlying low and high values of τ are instructive about how widely τ might vary, and why. Estimates of τ are quite low (0–0.1) for the p -beauty contest game when $p > 1$ and, consequently, the equilibrium is at the upper end of the range of possible choices (rows 1–2). In these games, subjects seem to have trouble realizing that they should choose very large numbers when $p > 1$ (though they equilibrate rapidly by learning; see Ho et al., 1998). Low τ ’s are also estimated among the PCC (Pasadena City College) subjects playing two- and three-player games (row 8 and 10). High values of τ (≈ 3 –5) appear in games where the equilibrium in the interior, 72 (row 7–10)—small incremental steps toward the equilibrium in these games produce high values of τ . High τ values are also estimated in games with an equilibrium of zero when subjects are professional stock market portfolio managers (row 19), Caltech students (row 20), game theorists (row 24), and subjects self-selecting to enter newspaper contests (row 25). The latter subject pools show that in high analytical and educated subject pools (especially with self-selection) τ can be much higher than in other subject pools.

A sensible intuition is that when stakes are higher, subjects will use more steps of reasoning (and may think others will think harder too). Rows 3 and 6 compare low stakes (\$1 per person per period) and high stakes (\$4) in games with an interior equilibrium of 72. When stakes are higher, τ is estimated to be twice as large (5.01 versus 2.51), which is a clue that some sort of cost-benefit analysis may underlie steps of reasoning (cf. Alaoui and Penta, 2013).

Notwithstanding these interesting outliers, there is also substantial regularity across very diverse subject pools and payoff steps. About half the samples have confidence intervals that include $\tau = 1.5$. Subsamples of corporate CEOs, high-functioning 80-year-old spouses of memory-impaired patients, and high school students (rows 13, 15–16) all have τ values from 1.1–1.7.

An interesting question about CH modeling is how persistent a player’s thinking step is across games. The few studies that have looked carefully found fairly stable estimated

steps within a subject across games of similar structure (Costa-Gomes et al., 2001; Stahl and Wilson, 1995). For example, Chong et al. (2004) estimated separate τ values for each of 22 one-shot games with mixed-equilibria. An assignment procedure then chose a most-likely step for each subject and each game (given the 22 τ estimates). For each subject, an average of their first 11 estimated steps and last 11 estimated steps were computed. The correlation coefficient of these two average steps was 0.61. This kind of within-subject stability is a little lower than many psychometric traits (e.g., intelligence, extraversion) and is comparable to econographic traits such as risk-aversion.

Two studies looked most carefully at within-subject step stability. Georgiadis et al. (2013) had subjects play variants of two structurally different games. They found modest stability of step choices within each game type, and low stability across the two game types. Hyndman et al. (2013) did the most thorough study; they measured both choices and beliefs across several baseline payoff-isomorphic games. They found that 30% of subjects appear to maintain a stable type belief (mostly step 1) across games, and others fluctuate between lower and higher belief between games. These studies bracket what we should expect to see about stability of step types (i.e., a mixture of stability). Since nothing in these theories commits to how stable “types” are, this empirical result is not surprising at all.

Finally, whether subjects maintain a single step type across games is neither predicted by the theory nor important for most applications. The most basic applications involve aggregate prediction, and sensitivity of predicted results to comparative static changes in game structure. It is very rare that the scientific goal is to predict what a specific subject will do in one game type, based on their behavior in a different game type.

10.2.2 Market entry games

Consider binary entry games in which there is a market with demand c (expressed as a percentage of the total number of potential entrants). Each of the many potential entrants decides simultaneously whether or not to enter the market. If a potential entrant thinks that fewer than $c\%$ will enter she will enter; if she thinks more than $c\%$ will enter she stays out.

There are three regularities in many experiments based on entry games like this one (see Camerer, 2003; Ochs, 1999; Seale and Rapoport, 2000, chapter 7): (1) Entry rates across different levels of demand c are closely correlated with entry rates predicted by (asymmetric) pure equilibria or symmetric mixed equilibria; (2) players slightly over-enter at low levels of demand and under-enter at high levels of demand; and (3) many players use noisy cutoff rules in which they stay out for levels of demand below some cutoff c^* and enter for higher levels of demand.

In Camerer et al. (2004), we apply the thinking model with best response (i.e., $\lambda = \infty$) to explain subject behaviors in this game. Step zero players randomize and enter

half the time. This means that when $c < 0.5$ one step thinkers stay out and when $c > 0.5$ they enter. Players doing two 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 two-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, one-step players are afraid to enter when $c < 1/2$. But when c is not too low (between $1/6$ and 0.5) the two-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 one-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 coordinate 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 second and third regularities discussed above for reasonable τ values. [Figure 10.1](#) plots entry rates from the first block of two studies for a game similar to the one above ([Seale and Rapoport, 2000](#); [Sundali et al., 1995](#)). 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 10.1](#) also shows the thinking steps entry function $N(all|\tau)(c)$ for $\tau = 1.25$. The function reproduces monotonicity and the over- and under-capacity effects. The thinking-steps models also produce approximate cutoff rule behavior for all higher thinking steps except two. When $\tau = 1.25$, step 0 types randomize, step 1 types enter for all c above 0.5 , step 2–4 types use cutoff rules with one “exception,” and steps 5–above use strict cutoff rules. This mixture of random, cutoff and near-cutoff rules is roughly what is observed in the data when individual patterns of entry across c are measured (e.g., [Seale and Rapoport, 2000](#)).

This example makes a crucial point about the goal and results from CH modeling. The goal is *not* (just) to explain nonequilibrium behavior. Another goal is to explain a *lack* of nonequilibrium behavior—that is, when is equilibration remarkably good, even with no special training or experience, and no opportunities for learning or communication? Note that in p -beauty contest games, if some players are out of equilibrium then even sophisticated players will prefer to be out of equilibrium. However, in entry games if some players over- or under-react to the capacity c then the sophisticated players will behave oppositely, leading to aggregate near-equilibrium. More generally, in games with strategic complementarity a little irrationality (e.g., step 0) will be multiplied; in games with strategic substitutes a little irrationality will be mitigated ([Camerer and Fehr, 2006](#)).

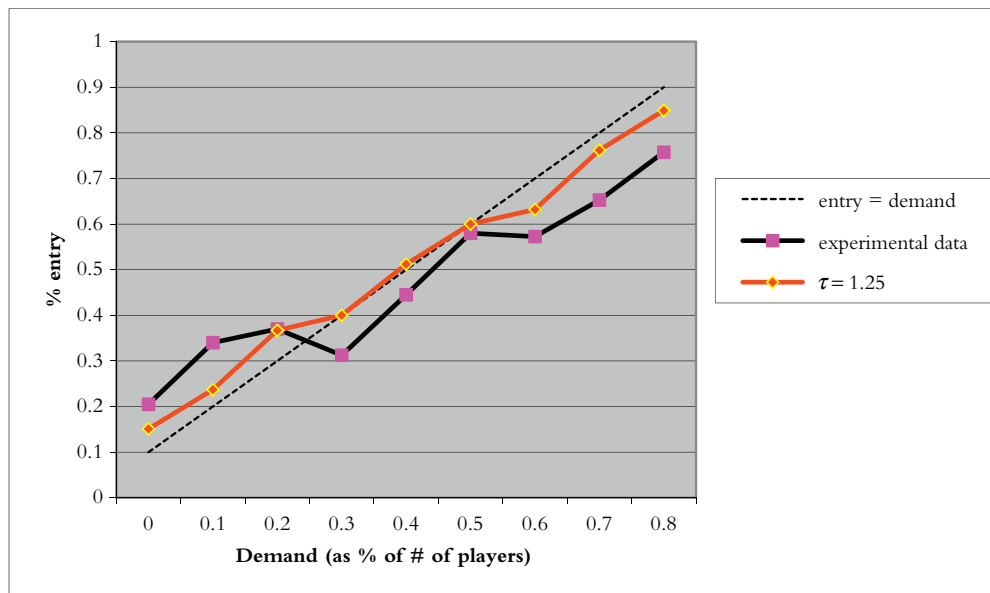


Figure 10.1 How entry varies with demand (D), experimental data and thinking model.

10.2.3 LUPI lottery

A unique field study used a simple lottery played in Sweden by an average of 53,783 players per day, over 7 weeks (Ostling et al., 2011). Participants in this lottery paid 1 euro to pick an integer from 1 to 99,999. The participant who chose the lowest unique positive integer won a prize of 10,000 euros (hence, the lottery is called LUPI). Interestingly, solving for the Nash equilibrium for a fixed number of n players is computationally intractable for large n . However, if the number of players is random and Poisson distributed across days, then the methods of Poisson games (Myerson, 1998) can be applied. The symmetric Poisson–Nash equilibrium (PNE) has an elegant simple structure where n is the mean number of players and p_k is the probability of playing integer k . The symmetric PNE has a striking nonlinear shape: numbers from 1 to around 5500 are chosen almost equally often, but with slightly declining probability (i.e., 1 is most likely, 2 is slightly less likely, etc.) (see Figure 10.2). A bold property of the PNE is that numbers above 5513—a range that includes 95% of all available numbers—should rarely be chosen.

Figure 10.3 shows the full distribution of number choices in the first week, along with the PNE and a best-fitting quantal CH model (i.e., λ is finite). Compared to the PNE, players chose too many low numbers, below 2000, and too many high numbers. As a result, not enough numbers between 2000 and 5000 are chosen. CH can fit these deviations from PNE with a value of $\tau = 1.80$, which is close to estimates from many experimental games. Intuitively, step 1 players choose low numbers because they think

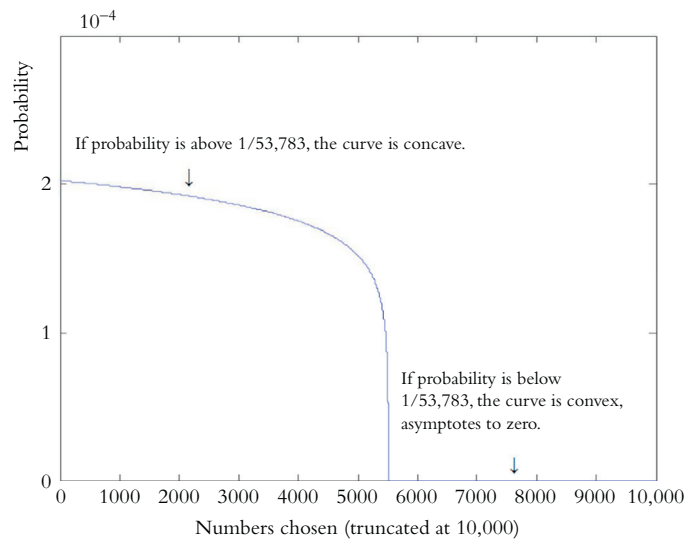


Figure 10.2 Poisson-Nash equilibrium for the LUP game (*the average number of players is 53,783*).

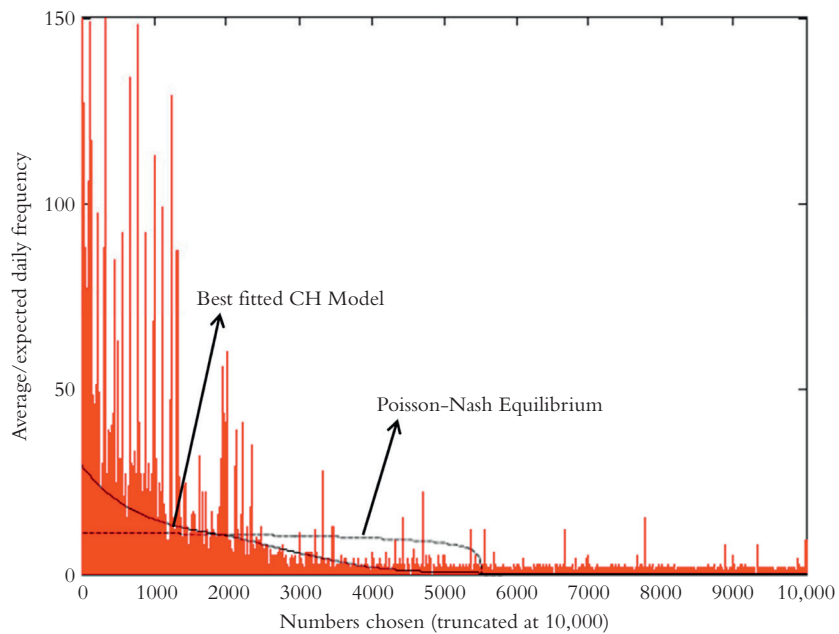


Figure 10.3 The full distribution of actual choices.

step 0 randomizers will pick too many high numbers, step 2 numbers choose above the step 1 choices, and so on.

Despite that CH can explain the deviations of the number choices in LUPI lottery, the actual behavior is not far from the PNE prediction in general, given how precisely bold the predicted distribution is. Recall that PNE predicts only numbers below 5513 will be played—excluding about 95% of the strategies—and overall, 93.3% of the numbers are indeed below that threshold. Furthermore, over seven weeks almost every statistical feature of the empirical distribution of numbers moved toward the PNE. For example, PNE predicts the average number will be 2595. The actual averages were 4512 in the first week and 2484 in the last week (within 4% of the prediction). A scaled-down laboratory version of LUPI also replicated these basic patterns.

10.2.4 Summary

Simple models of thinking steps attempt to predict choices in one-shot games and provide initial conditions for learning models. CH and level-k approaches incorporate discrete steps of iterated thinking. In Poisson-CH, the frequencies of players using different numbers of steps is Poisson-distributed with mean τ . While these models have been applied to hundreds of experimental data sets, and several field settings, in this chapter we focussed on data from dominance-solvable p -beauty contests, market entry, a LUPI field lottery, and a private-information auction (see [Section 10.3](#)). For Poisson-CH, estimates of τ are typically around 1–2. Importantly, CH can typically explain large deviations from equilibrium *and* surprising cases in which equilibration occurs without communication or experience (in entry games).

10.3. QUANTAL RESPONSE EQUILIBRIUM

The CH and LK models explain nonequilibrium behavior by mistakes in beliefs about what other subjects will do, due to bounded rationality in strategic thinking. A different approach is to assume that beliefs are accurate, but responses are noisy. This approach is called QRE ([McKelvey and Palfrey, 1995](#)). QRE relaxes the assumption that players always choose the best actions given their beliefs by incorporating “noisy” or “stochastic” response. The theory builds in a sensible principle that actions with higher expected payoffs are chosen more often; that is, players “better-respond” rather than “best-respond.” If the response function is logit, QRE is defined by

$$\begin{aligned}
 A_i^j(0|K) &= \sum_{h=1}^{m_i} \pi_i(s_i^j, s_{-i}^h) \cdot P_{-i}^h(1) \\
 P_i^j(1) &= \frac{e^{\lambda \cdot A_i^j(0)}}{\sum_{h=1}^{m_i} e^{\lambda \cdot A_i^h(0)}}
 \end{aligned} \tag{10.4}$$

Mathematically, the QRE nests Nash equilibrium as a special case. Specifically, when the noise parameter λ goes to infinity, QRE converges to Nash equilibrium.

The errors in the players' QRE best-response functions are usually interpreted as decision errors in the face of complex situations or as unobserved latent disturbances to the players' payoffs (i.e., the players are optimizing given their payoffs, but there is a component of their payoffs that only they understand). In other words, the relationship between QRE and Nash equilibrium is analogous to the relationship between stochastic choice and deterministic choice models. Note that QRE is different from trembling hand perfect equilibrium in that the noise parameter λ in QRE is part of the equilibrating process and players use it to determine an action's expected payoff as well as its choice probability.

QRE has been used successfully in many applications to explain deviations from Nash equilibrium in games (Goeree and Holt, 2001; McKelvey and Palfrey, 1995; Rosenthal, 1981).⁵ The key feature of the approach is that small mistakes can occur, but if a small mistake by player i has a large impact on player j , QRE prediction can be far from Nash equilibrium. QRE can also be used as a tool to check robustness in designing institutions or predicting how structural changes will change behavior.

10.3.1 Asymmetric hide-and-seek game

Let's see how QRE can improve equilibrium prediction by considering a game with an unique mixed-strategy equilibrium (see Table 10.2 for its payoff matrix). The row player's strategy space consists of actions A1 and A2, while the column player's chooses between B1 and B2. The game is a model of "hide-and-seek" in which one player wants to match another player's numerical choice (e.g., A1 responding to B1), and another player wants to mismatch (e.g., B1 responding to A2). The row player earns either 9 or 1 from matching on (A1, B1) or (A2, B2) respectively. The column player earns 1 from mismatching on (A1, B2) or (A2, B1).

The empirical frequencies of each of the possible actions, averaged across many periods of an experiment conducted on this game, are also shown in Table 10.2 (McKelvey and Palfrey, 1995). What is the Nash equilibrium prediction for this game? We start by observing that there is no pure-strategy Nash equilibrium for this game so we look for a mixed-strategy Nash equilibrium. Let us suppose that the Row player chooses A1 with probability p and A2 with probability $1 - p$, and the Column player chooses B1 with probability q and B2 with probability $1 - q$. In a mixed-strategy equilibrium, the players actually play a probabilistic mixture of the two strategies. If their valuation of

⁵ To use QRE to predict behaviors ex ante, one must know the noise parameter (λ). This can be accomplished by first estimating λ in a similar game. Previous studies appear to show that λ does vary from game to game. Hence, an open research question is to develop a theory that maps features of a game to the value of λ so that one can then use QRE to predict behavior in any new game.

Table 10.2 Asymmetric hide-and-seek game

	B1 q	B2 $1-q$	Empirical frequency ($N = 128$)	Nash equilibrium	QRE
A1 p	9, 0	0, 1	0.54	0.50	0.65
A2 $1-p$	0, 1	1, 0	0.46	0.50	0.35
Empirical Frequency	0.33	0.67			
Nash Equilibrium	0.10	0.90			
QRE	0.35	0.65			

outcomes is consistent with expected utility theory, they only prefer playing a mixture if they are indifferent between each of their pure strategies. This property gives a way to compute the equilibrium mixture probabilities p and q . The mixed-strategy Nash equilibrium for this game turns out to be $[(0.5A1, 0.5A2), (0.1B1, 0.9B2)]$. Comparing this with the empirical frequencies, we find that Nash prediction is close to actual behavior by the Row players, whereas it under-predicts the choice of B1 for the Column players.

If one player plays a strategy that deviates from the prescribed equilibrium strategy, then according to the optimization assumption in Nash equilibrium, the other player must best-respond and deviate from Nash equilibrium as well. In this case, even though the predicted Nash equilibrium and actual empirical frequencies almost coincide for the Row player, the players are not playing a Nash equilibrium jointly, because the Row player should have played differently given that the Column player has deviated quite far from the mixed-strategy Nash equilibrium (playing B1 33% of the time rather than 10%).

We now illustrate how to derive the QRE for a given value of λ in this game. Again, suppose the Row player chooses A1 with probability p and A2 with probability $1 - p$, while the Column player chooses B1 with probability q and B2 with probability $1 - q$. Then the expected payoffs from playing A1 and A2 are $q * 9 + (1 - q) * 0 = 9q$ and $q * 0 + (1 - q) * 1 = 1 - q$ respectively. Therefore, we have $p = \frac{e^{\lambda \cdot 9q}}{e^{\lambda \cdot 9q} + e^{\lambda \cdot (1-q)}}$. Similarly, the expected payoffs to B1 and B2 for the Column player are $1 - p$ and p respectively, so we have $q = \frac{e^{\lambda \cdot (1-p)}}{e^{\lambda \cdot (1-p)} + e^{\lambda \cdot p}}$. Notice that q is on the right hand side of the first equation, which determines p , and p is on the right hand side of the second equation, which determines q . For any value of λ , there is only one pair of (p, q) values that solves the simultaneous equations and yields a QRE. If $\lambda = 2$, for example, the QRE predictions are $p^* = 0.646$ and $q^* = 0.343$ which are closer to the empirical frequencies than the Nash equilibrium predictions are. If $\lambda = \infty$, the QRE predictions are $p^* = 0.5$ and $q^* = 0.1$, which are identical to Nash equilibrium predictions.

Using the actual data, a precise value of λ can be estimated using maximum-likelihood methods. The estimated λ for the QRE model for the asymmetric

“hide-and-seek” game is 1.95. The negative of the log likelihood of QRE (an overall measure of goodness of fit) is 1721, a substantial improvement over a random model benchmark ($p = q = 0.5$) which has a fit of 1774. The Nash equilibrium prediction has a fit of 1938, which is worse than random (because of the extreme prediction of $q = 0.1$).

10.3.2 Maximum value auction

We describe a private-information auction which illustrates another application of QRE. In the maximum-value second price auction, two players observe private integer signals x_1 and x_2 drawn independently from a commonly known distribution in the interval $[0, 10]$ (Bulow and Klemperer, 2002; Carrillo and Palfrey, 2011). They bid for an object which has a common value equal to the maximum of those two signals, $\max(x_1, x_2)$. The highest bidder wins the auction and pays a price equal to the second-highest bid.

How should you bid? Bidding less than your signal is a mistake because your bid just determines whether you win, not what you pay; so bidding less never saves money. In fact, you might miss a chance to win at a price below the object’s value if you bid too low and get outbid, so underbidding is weakly dominated. Second, bidding above your signal could be a mistake: If the other bidder is also overbidding, either of you may get stuck overpaying for an item with a low maximum-value. In the unique symmetric (weak) BNE, therefore, players simply bid their values. In fact, the symmetric equilibrium where both players bid their signal can be solved for with two rounds of elimination of weakly dominated strategies.

Note that the unique symmetric BNE in which bidders bid their signals, but the equilibrium is weak; indeed it is just about as weak as an equilibrium can possibly be. If the other bidder is bidding according to the Nash equilibrium, i.e., bidding their signals, then *every bid greater than or equal to your signal* is a (weak) best response to Nash equilibrium.

QRE impose the assumption that beliefs about choices of other players are accurate, but allow imperfect (noisy) response. A natural conjecture is that (symmetric) regular QRE (Goeree et al., 2005) in these auctions will typically entail frequent overbidding, because bidding a little too high is a very small mistake.

In these experiments (Ivanov et al., 2010) subjects first participated in 11 independent auctions, once for each private signal 0–10, *with no feedback*. For each possible value, the rectangles in Figures 10.4ab show the median bid (thick horizontal line), and the vertical boundaries of the rectangle are the first and third quartiles. The main result is that there is substantial overbidding which has a “hockey stick” shape: Bids when signals are below \$5 are around \$4–\$5 and do not change much as the signal increases. Bids when signals are above \$5 are a little above the signals, and increase as signals increase. The BNE is just to bid the signal value (shown by a dotted 45-degree line). To fit the data, hierarchical QRE (HQRE) and CH-QR (with quantal response and only levels 0–2) were first fit to

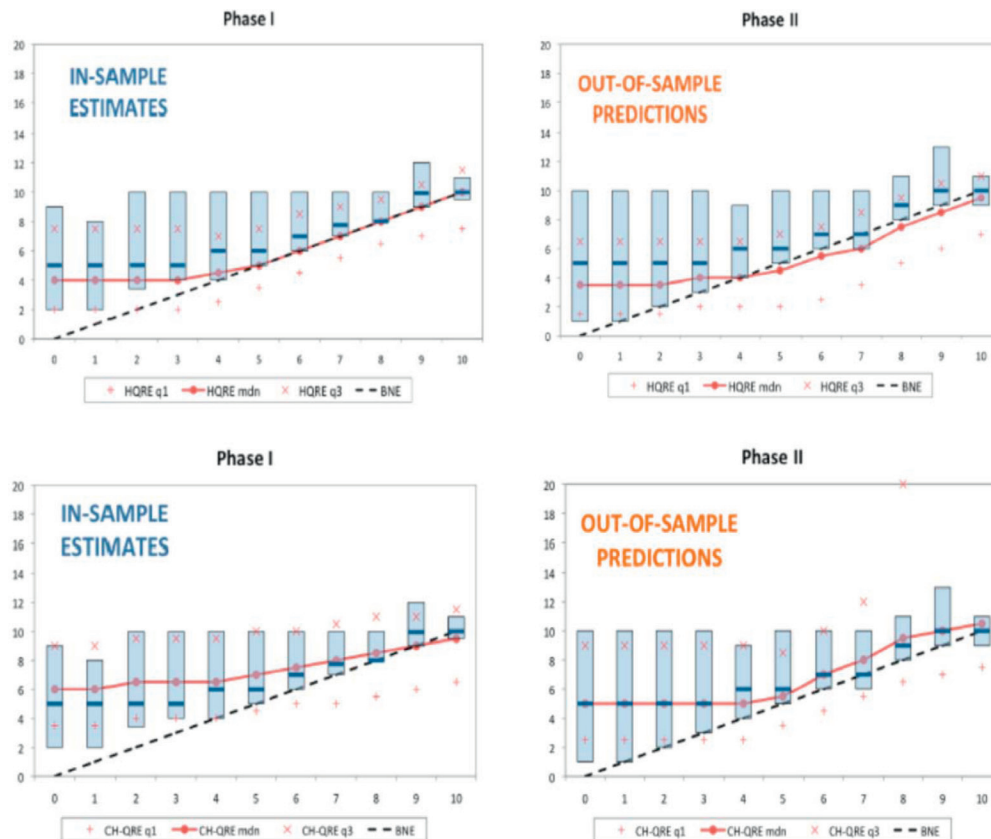


Figure 10.4 Bids by Signal, MinBid Treatment. Data from [Nunnari et al. \(2012\)](#).

the bidding data from phase I. Those best-fitting parameters were then used to predict bids in phase II. Note that HQRE and CH-QR make probabilistic predictions—that is, they predict the distribution of likely bids given each signal. The predictions are therefore represented by a dot for the predicted median and small x's and +s for the first and third quartiles. Visually, a good fit is indicated if, for each signal, the dot is near the horizontal line and the x's and +s are near the rectangle boundaries.

Both models do a reasonable job of picking up the major regularity in how bids deviate from BNE, which is the hockey stick pattern. However, HQRE predicts bids which are a little too low (including a substantial rate of *underbidding*). CH-QR fits very nicely, except at the highest signals \$8–10, where it predicts too many high bids.

In this case, the essential feature of QRE approaches is that small mistakes are expected to occur, but can predict bids far from BNE. Since low signal values are unlikely to determine the object's value (the other signal is likely to be the maximum), there is only

a small loss from bidding too high on low signals. However, with high signals the bidder's own signal is likely to be the maximum, so overbidding is a bigger mistake and should be less common. These two intuitions generate the hockey stick shape that is evident in the data.

This example illustrates how different best-response equilibrium analysis can be from either CH or QRE. Curiously, [Ivanov et al. \(2010\)](#) conclude from their data that “Overall, our study casts a serious doubt on theories that posit the WC [winner's curse] is driven by beliefs [i.e., by CH and cursed equilibrium].” (p. 1435) However, as [Figure 10.4b](#) shows, the nonequilibrium belief theory CH fits the data quite well.

Finally, recall the LUP lottery game described previously, in which the player with a lowest unique positive integer wins a fixed prize. It happens that LUP is well suited for distinguishing CH and QRE approaches. For LUP games with a small number of players, for which both QRE and CH can be solved numerically, the QRE distribution approaches NE from below, while the CH distribution predicts more low number choices than NE. Thus, QRE cannot explain the large number of low number choices occurring in the data, as CH can.

However, in a comparison of several other games, we have typically found that QRE and CH make about equally accurate predictions (e.g., [Rogers et al., 2009](#)). [Moinas and Pouget \(2013\)](#) use a remarkable bubble investment game to carefully compare models and find good fit from a heterogeneous form of QRE. A careful meta-analysis by [Wright and Leyton-Brown \(2013\)](#) shows that Poisson-CH generally fits a little better than other versions (given its parsimony) and all limited-thinking approaches do better than equilibrium in predicting behavior.

10.4. 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.⁶ The 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 wide range of classes of games.

This section describes a one-parameter theory of learning in decisions and games called functional EWA (or fEWA for short; also called “EWA Lite” to emphasize its ‘low-calorie’ parsimony) ([Ho et al., 2007](#)). fEWA predicts the time path of individual behavior in any normal-form game given the initial conditions. Initial conditions can be imposed or estimated in various ways. We use the CH model described in the previous

⁶ To name only a few examples, see [Camerer \(1987\)](#) (partial adjustment models); [Smith et al. \(1988\)](#) and [Camerer and Weigelt \(1993\)](#) (Walrasian excess demand); [McAllister \(1991\)](#) (reinforcement); [Roth and Erev \(1995\)](#) (reinforcement); [Ho and Weigelt \(1996\)](#) (reinforcement and belief learning); [Cachon and Camerer \(1996\)](#) (Cournot dynamics).

section to specify the initial conditions. The goal is to predict both initial conditions and equilibration in new games in which behavior has never been observed, with minimal free parameters (the model uses two, τ and λ).

10.4.1 Parametric EWA learning: interpretation, uses, and limits

fEWA is a relative of a 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 best-response CH model is used to generate initial attractions given parameter value τ .

Denote i 's j 'th strategy by s_i^j , chosen strategies by i and other players (denoted $-i$) by $s_i(t)$ and $s_{-i}(t)$, and player i 's payoffs by $\pi_i(s_i^j, s_{-i}(t))$.⁷ 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

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{N(t-1)\phi(1-\kappa) + 1} \quad [10.5]$$

and the experience weight (the “EW” part) is updated according to $N(t) = N(t-1)\phi(1-\kappa) + 1$.

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 foregone payoffs from unchosen strategies ($s_i^j \neq s_i(t)$) is δ .

Attractions are mapped into choice probabilities using a logit response function given by:

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}} \quad [10.6]$$

where λ is the response sensitivity.⁸ 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.⁹

⁷ To avoid complications with negative payoffs, we rescale payoffs by subtracting them by the minimum payoff so that rescaled payoffs are always weakly positive.

⁸ Note that we can use the same parameter λ in both [10.3] and [10.6]. The parameter maps attractions into probabilities.

⁹ Other models aim to explain aggregated choices at the population level. Of course, models of this sort can sometimes be useful. But our view is that a parsimonious model which can explain very fine-grained data can probably explain aggregated data well too.

Each EWA parameter has a natural interpretation:

- The parameter δ is the weight placed on foregone payoffs. It presumably 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 foregone payoffs.
- The parameter ϕ decays previous attractions due to 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 κ controls the rate at which attractions grow. When $\kappa = 0$ attractions are weighted averages and grow slowly; when $\kappa = 1$ 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 κ), and “exploiting” what is known by locking in to a good strategy (high κ) (e.g., Sutton and Barto, 1998).
- The initial experience weight $N(0)$ is like a strength of prior beliefs in models of Bayesian belief learning. It plays a minimal empirical role so it is set to one 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 what strategies they choose (see Brown, 1951; Fudenberg and Levine, 1998). EWA shows that reinforcement and belief learning, which were often treated as fundamentally different, are actually related in a non-obvious way, because both are special kinds of a general reinforcement rule¹⁰. When $\delta = 0$ the EWA rule is a simple reinforcement rule.¹¹ When $\delta = 1$ and $\kappa = 0$ the EWA rule is equivalent to belief learning using weighted fictitious play.¹²

Foregone 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 Stoecker, 1986). Suppose players follow EWA but don’t know foregone payoffs, and believe those payoffs are monotonically

¹⁰ See also Cheung and Friedman (1997, pp. 54–55), Fudenberg and Levine (1998, pp. 184–185); and Hopkins (2002).

¹¹ See Bush and Mosteller (1955), Harley (1981), Cross (1983), Arthur (1991), McAllister (1991), Roth and Erev (1995), and Erev and Roth (1998).

¹² 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 foregone). The beliefs themselves are an epiphenomenon that disappear when the updating equation is written expected payoffs rather than beliefs.

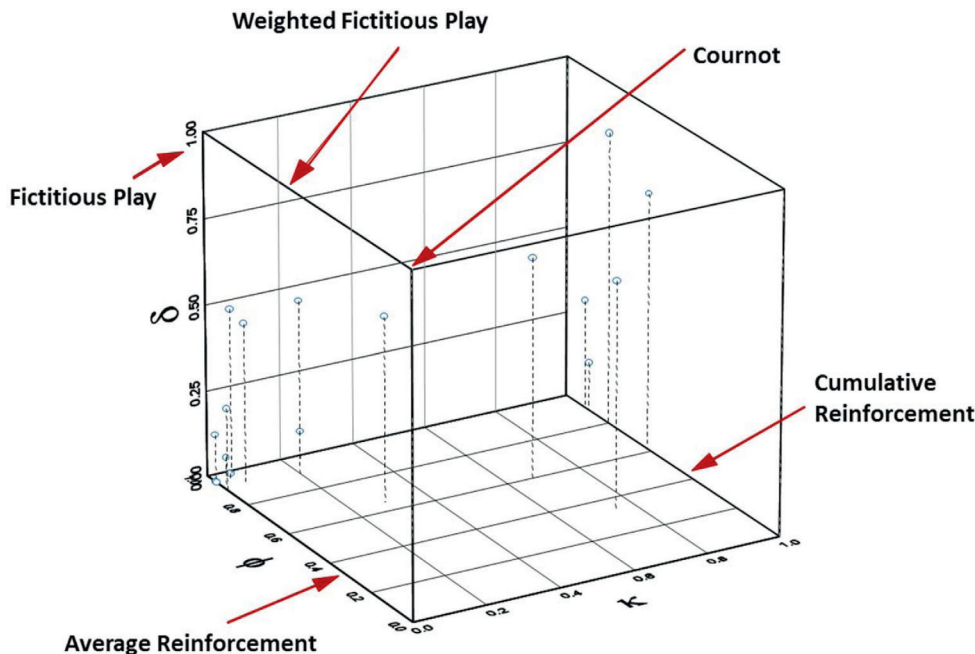


Figure 10.5 EWA cube.

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 behavior will look like direction learning. Imitating a player who is similar and successful can also be seen as a way of heuristically inferring high foregone payoffs from an observed choice and moving in the direction of those higher payoffs.

The relation of various learning rules can be shown visually in a cube showing configurations of parameter values (see Figure 10.5). Each point in the cube is a triple of EWA parameter values which specifies a precise updating equation. The corner of the cube with $\phi = \kappa = 0, \delta = 1$, is Cournot best-response dynamics. The corner $\kappa = 0, \phi = \delta = 1$, is standard fictitious play. The vertex connecting these corners, $\delta = 1, \kappa = 0$, is the class of weighted fictitious play rules (e.g., Fudenberg and Levine, 1998). The vertices with $\delta = 0$ and $\kappa = 0$ or 1 are averaging and cumulative choice reinforcement rules (Erev and Roth, 1998; Roth and Erev, 1995).

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 than the special cases of reinforcement and weighted fictitious in most cases, except in games with mixed-strategy equilibrium where reinforcement does equally well.¹³ In our

¹³ In mixed games no model improves much on Nash equilibrium (and often don't improve on QRE at all), and parameter identification is poor; see Salmon, 2001.

model estimation and validation, we always penalize the EWA model in ways that are known to make the *adjusted fit worse* if a model is too complex (i.e., if the data are actually generated by a simpler model).¹⁴ Furthermore, econometric studies show that if the data were generated by simpler belief or reinforcement models, then EWA estimates would correctly identify that fact for most games and reasonable sample sizes (see [Cabrales and Garcia-Fontes, 2000](#); [Salmon, 2001](#)). Since EWA is capable of identifying behavior consistent with special cases, when it does not then the hybrid parameter values are improving fit.

[Figure 10.5](#) 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 κ . 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 behavior in brand new games.

10.4.2 fEWA functions

This concern is addressed by a version of EWA, fEWA, which replaces free parameters with specific values and functions that 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 fEWA has only one parameter which must be estimated (λ),¹⁵ 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.

As shown in [Figure 10.5](#), the front face of the cube ($\kappa = 0$) captures almost all familiar special cases except for the cumulative reinforcement model. The cumulative model has been supplanted by the averaging model (with $\kappa = 0$) because the latter seems to be more robust in predicting behavior in some games (see [Roth and Erev, 1995](#)). This sub-class of EWA learning models is the simplest model that nests averaging reinforcement and weighted fictitious play (and hence Cournot and simple fictitious play) as special cases.

¹⁴ We typically 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.

¹⁵ Note that if your statistical objective is to maximize hit rate, λ does not matter and so fEWA is a zero-parameter theory given initial conditions.

It can also capture a weighted fictitious play model using time-varying belief weights (such as the stated-beliefs model explored by Nyarko and Schotter, 2002), as long as subjects are allowed to use a different weight to decay lagged attractions over time (i.e., move along the top edge of the side in Figure 10.5). There is also an empirical reason to set κ to a particular value. Our prior work suggests that κ does not seem to affect fit much (e.g., Ho et al., 2008). The initial experience $N(0)$ was included in the original EWA model so that Bayesian learning models are nested as a special case— $N(0)$ represents the strength of prior beliefs. We restrict $N(0) = 1$ here because its influence fades rapidly as an experiment progresses and most subjects come to experiments with weak priors anyway.

Consequently, we are left with three free parameters— ϕ , δ , and λ . To make the model simple to estimate statistically, and self-tuning, the parameters ϕ and δ are replaced by deterministic functions $\phi_i(t)$ and $\delta_{ij}(t)$ of player i 's experience with strategy j , up to period t . These functions determine numerical parameter values for each player, strategy, and period, which are then plugged into the EWA updating equation above to determine attractions in each period. Updated attractions determine choice probabilities according to the logit rule, given a value of λ . Standard maximum-likelihood methods for optimizing fit can then be used to find which λ fits best.¹⁶

10.4.2.1 The change-detector function $\phi_i(t)$

The decay rate ϕ which weights lagged attractions is sometimes called “forgetting” (an interpretation, which is carried over from reinforcement models of animal learning). While forgetting obviously does occur, the more interesting variation in $\phi_i(t)$ across games, and across time within a game, is a player's perception of how quickly the learning environment is changing. The function $\phi_i(t)$ should therefore “detect change.” When a player senses that other players are changing, a self-tuning $\phi_i(t)$ should dip down, putting less weight on distant experience. As in physical change detectors (e.g., security systems or smoke alarms), the challenge is to detect change when it is really occurring, but not falsely mistake small fluctuations for real changes too often.

The core of the $\phi_i(t)$ change-detector function is a “surprise index,” which is the difference between other players' most recently chosen strategies and their chosen strategies in all previous periods. First define a cumulative history vector, across the other players' strategies k , which records the historical frequencies (including the last period t) of the choices by other players. The vector element $h_i^k(t)$ is $\frac{\sum_{\tau=1}^t I(s_{-i}^k - s_{-i}(\tau))}{t}$.¹⁷

¹⁶ If one is interested only in the hit rate—the frequency with which the predicted choice is the same as what a player actually picked—then it is not necessary to estimate λ . The strategy that has the highest attraction will be the predicted choice. The response sensitivity λ only dictates how frequently the highest-attraction choice is actually picked, which is irrelevant if the statistical criterion is hit rate.

¹⁷ Note that if there is more than one other player, and the distinct choices by different other player's matter to player i , then the vector is an $n - 1$ -dimensional matrix if there are n players.

The immediate ‘history’ $r_i^k(t)$ is a vector of 0’s and 1’s which has a one for strategy $s_{-i}^k = s_{-i}(t)$ and 0’s for all other strategies s_{-i}^k (i.e., $r_i^k(t) = I(s_{-i}^k, s_{-i}(t))$). The surprise index $S_i(t)$ simply sums up the squared deviations between the cumulative history vector $h_i^k(t)$ and the immediate history vector $r_i^k(t)$; that is,

$$S_i(t) = \sum_{k=1}^{m-i} (h_i^k(t) - r_i^k(t))^2. \quad [10.7]$$

In other words, the surprise index captures the degree of change of the most recent observation¹⁸ from the historical average. Note that it 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 ‘forever’ and suddenly switches to something brand new). When surprise index is zero, we have a stationary environment; when it is two, we have a turbulent environment. The change-detecting decay rate is:

$$\phi_i(t) = 1 - \frac{1}{2} \cdot S_i(t). \quad [10.8]$$

Because $S_i(t)$ is between zero and two, ϕ is always (weakly) between one and zero.

Some numerical boundary cases help illuminate how the change-detection works. If the other player chooses the strategy she 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). The opposite case is when an opponent has previously chosen a single strategy in every period, and suddenly switches to a new strategy. In that case, $\phi_i(t)$ is $\frac{2t-1}{t^2}$. This expression declines gracefully toward 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 bigger surprise (and should have an associated lower ϕ) if it follows a *longer* string of *identical* choices which are different from the surprising new choice. Note that since the observed behavior in period t is included in the cumulative history $h_i^k(t)$, $\phi_i(t)$ will never dip completely to zero. (which could be a mistake because it erases all the history embodied in the lagged attraction). For example, if a player chose the same strategy for each of 19 periods and a new strategy in period 20, then $\phi_i(t) = 39/400 = 0.0975$.

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

¹⁸ 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 which are played with positive probability, the surprise index defines recent history over a window of the last W periods (e.g., in a game with four strategies that are played in equilibrium, $W = 4$). Then $r_i^k(t) = \sum_{k=1}^{m-i} \left[\frac{\sum_{\tau=t-W+1}^t I(s_{-i}^k, s_{-i}(\tau))}{W} \right]$.

and asymptotes at 0.5 as t increases. Comparing the case where the previous strategy was the same, and the previous strategies were all different, it is evident that if the choice in period t is new, the value of $\phi_i(t)$ is higher if there were more variation in previous choices, and lower if there were less variation. This mimics a hypothesis-testing approach in which more variation in previous strategies implies that players are less likely to conclude there has been a regime shift, and therefore do not lower the value of $\phi_i(t)$ too much.

Note that the change-detector function $\phi_i(t)$ is individual and time specific and it depends on information feedback. Nyarko and Schotter (2002) show that a weighted fictitious play model that uses stated beliefs (instead of empirical beliefs posited by the fictitious play rule) can predict behavior better than the original EWA model in games with unique mixed-strategy equilibrium. One way to interpret their result is that their model allows each subject to attach a different weight to previous experiences over time. In the same vein, the proposed change-detector function allows for individual and time heterogeneity by positing them theoretically.

10.4.2.2 The attention function, $\delta_{ij}(t)$

The parameter δ is the weight on foregone payoffs. Presumably this is tied to the attention subjects pay to alternative payoffs, ex-post. Subjects who have limited attention are likely to focus on strategies that would have given higher payoffs than what was actually received, because these strategies present missed opportunities, which show that such a regret-driven rule converges to correlated equilibrium (Hart and Mas-Colell, 2001). To capture this property, define¹⁹

$$\delta_{ij}(t) = \begin{cases} 1 & \text{if } \pi_i(s_i^j, s_{-i}(t)) \geq \pi_i(t), \\ 0 & \text{otherwise.} \end{cases} \quad [10.9]$$

That is, subjects reinforce chosen strategies and all unchosen strategies with (weakly) better payoffs by a weight of one. They reinforce unchosen strategies with strictly worse payoffs by zero.

Note that this $\delta_{ij}(t)$ can transform the self-tuning rule into special cases over time. If subjects are strictly best-responding (ex post), then no other strategies have a higher ex-post payoff. Hence $\delta_{ij}(t) = 0$ for all strategies j which were not chosen, reducing the model to choice reinforcement. However if they always choose the worst strategy, then $\delta_{ij}(t) = 1$, which corresponds to weighted fictitious play. If subjects neither choose

¹⁹ In games with unique mixed-strategy equilibrium, we use $\delta_{ij}(t) = \frac{1}{W}$ if $\pi_i(s_i^j, s_{-i}(t)) \geq \pi_i(t)$ and 0 otherwise. This modification is driven by the empirical observation that estimated δ 's are often close to zero in mixed games (which might also be due to misspecified heterogeneity). Using only $\delta_{ij}(t)$ without this adjustment produces slightly worse fits in the two mixed-equilibrium games examined below where the adjustment matters (patent-rate games and the Mookerjee-Sopher games).

the best nor the worst strategy, the updating scheme will push them (probabilistically) towards those strategies that yield better payoffs, as is both characteristic of human learning and normatively sensible.

The updating rule is a natural way to formalize and extend the “learning direction” theory of [Selten and Stoecker \(1986\)](#). Their theory consists of an appealing property of learning: Subject move in the direction of ex-post best-response. Broad applicability of the theory has been hindered by defining “direction” only in terms of numerical properties of ordered strategies (e.g., choosing ‘higher prices’ if the ex-post best response is a higher price than the chosen price). The self-tuning $\delta_{ij}(t)$ defines the “direction” of learning set-theoretically, as shifting probability toward the set of strategies with higher payoffs than the chosen ones.

The self-tuning $\delta_{ij}(t)$ also creates the “exploration-exploitation” shift in machine learning (familiar to economists from multi-armed bandit problems). In low-information environments, it makes sense to explore a wide range of strategies, then gradually lock in to a choice with a good historical relative payoffs. In self-tuning EWA, if subjects start out with a poor choice, many unchosen strategies will be reinforced by their (higher) foregone payoffs, which shifts choice probability to those choices and captures why subjects “explore.” As equilibration occurs, only the chosen strategy will be reinforced, thereby producing an “exploitation” or “lock-in.” This is behaviorally very plausible. The updating scheme also helps to detect any change in environment. If a previously optimal response becomes inferior because of an exogenous change, other strategies will have higher ex-post payoffs, triggering higher $\delta_{ij}(t)$ values (and reinforcement of superior payoffs) and guiding players to re-explore better strategies.

The self-tuning $\delta_{ij}(t)$ function can also be seen as a reasonable all-purpose rule which conserves a scarce cognitive resource—attention. The parametric EWA model shows that weighted fictitious play is equivalent to generalized reinforcement in which all strategies are reinforced. But reinforcing many strategies takes attention. As equilibration occurs, the set of strategies which receive positive $\delta_{ij}(t)$ weights shrinks so attention is conserved when spreading attention widely is no longer useful. When an opponent’s play changes suddenly, the self-tuning $\phi_i(t)$ value drops. This change reduces attractions (since lagged attractions are strongly decayed) and spreads choice probability over a wider range of strategies due to the logit response rule. This implies that the strategy chosen may no longer be optimal, leading $\delta_{ij}(t)$ to allocate attention over a wider range of better-responses. Thus, the self-tuning system can be seen as procedurally rational (in Herbert Simon’s language) because it follows a precise algorithm and is designed to express the basic features of how people learn—by exploring a wide range of options, locking in when a good strategy is found, but re-allocating attention when environmental change demands such action.

A theorist’s instinct is to derive conditions when flexible learning rules choose parameters optimally, which is certainly a direction to explore in future research.

However, broadly optimal rules will likely depend on the set of games an all-purpose learning agent encounters, and also may depend sensitively on how cognitive costs are specified (and should also jibe with data on the details of neural mechanisms, which are not yet well-understood). So it is unlikely to find a universally optimal rule that can always beat rules which adapt locally.

Our approach is more like the exploratory work in machine learning. Machine learning theorists try to develop robust heuristic algorithms which learn effectively in a wide variety of low-information environments. Good machine learning rules are not provably optimal but perform well on tricky test cases and natural problems like those which good computerized robots need to perform (navigating around obstacles, hill-climbing on rugged landscapes, difficult pattern recognition, and so forth).

Before proceeding to estimation, it is useful to summarize the properties of the self-tuning model. First, the use of simple fictitious play and reinforcement theories in empirical analysis is often justified by the fact that they have only a few free parameters. The self-tuning EWA is useful by this criterion because it requires estimating only one parameter, λ (which is difficult to do without in empirical work). Second, the functions in self-tuning EWA naturally vary across time, people, games, and strategies. The potential advantage of this flexibility is that the model can predict across new games better than parametric methods. Whether this advantage is realized will be examined below. Third, the self-tuning parameters can endogenously shift across rules. Early in a game, when opponent choices vary a lot and players are likely to make ex-post mistakes, the model automatically generates low values of $\phi_i(t)$ and high $\delta_{ij}(t)$ weights—it resembles Cournot belief learning. As equilibration occurs and behavior of other players stabilizes, $\phi_i(t)$ rises and $\delta_{ij}(t)$ falls—it resembles reinforcement learning. The model therefore keeps a short window of history (low ϕ) and pays a lot of attention (high δ) when it should, early in a game, and conserves those cognitive resources by remembering more (high ϕ) and attending to fewer foregone strategies (low δ) when it can afford to, as equilibration occurs.

fEWA has three advantages. First, it is easy to use because it has only one free parameter (λ). Second, parameters in fEWA 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 (e.g., [Salmon, 1999](#), and [Stahl, 1996, 2000](#)). 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 behavior of fEWA than about some parametric models. A key feature of fEWA is that as a player’s opponents’ behavior stabilizes, $\phi_i(t)$ goes toward one and (in games with pure equilibria) $\delta_i(t)$ does too. Since $\kappa = 0$, fEWA automatically turns into fictitious play; and a lot is known about theoretical properties of fictitious play.

10.4.3 fEWA predictions

In this section we compare in-sample fit and out-of-sample predictive accuracy of different learning models when parameters are freely estimated, and check whether fEWA 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 patent race games (Rapoport and Amaldoss, 2000); a median-action order statistic coordination game with several players (Van Huyck et al., 1990); a continental-divide coordination game, in which convergence behavior is extremely sensitive to initial conditions (Van Huyck et al., 1997); dominance-solvable p-beauty contests (Ho et al., 1998); and a price-matching game (called “travelers’ dilemma” by Capra et al., 1999).

The estimation procedure for fEWA is sketched briefly here (see Ho et al., 2007 for details). Consider a game where N subjects play T rounds. For a given player i , 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:

$$\prod_{t=1}^T P_i^{s_i(t)}(t) \quad [10.10]$$

The joint likelihood function L of observing all players’ choice is given by

$$L(\lambda) = \prod_i^N [\prod_{t=1}^T P_i^{s_i(t)}(t)] \quad [10.11]$$

Most models use first-period data to determine initial attractions. We also compare all models with burned-in attractions with a model in which the thinking steps model from the previous section is used to create initial conditions and combined with fEWA. 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,²⁰ attractions are updated using the EWA formula. fEWA 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% of the subjects. Then the value of λ is frozen and used to forecast behavior of the entire path of the remaining 30% of the subjects. Payoffs were all converted to dollars (which is important for cross-game forecasting).

In addition to fEWA (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; Erev et al., 2002), and QRE.

²⁰ The initial parameter values are $\phi_i(0) = 0.5$ 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 $\frac{t-1}{t}$.

The first question we ask is how well models fit and predict on a game-by-game basis (i.e., parameters are estimated separately for each game). For out-of-sample validation we report both hit rates (the fraction of most-likely choices which are picked) and log likelihood (LL). Recall that these results forecast a holdout sample of subjects *after* model parameters have been estimated on an earlier sample and then “frozen” for holdout. If a complex model is fitting better within a sample purely because of spurious overfitting, it will predict more poorly out of sample. Results are summarized in [Table 10.3](#).

The best fits for each game and criterion are printed in bold; hit rates that are statistically indistinguishable from the best (by the McNemar test) are also in bold. 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. fEWA also does well on hit rate in six 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 worst.

Combining fEWA with CH model to predict initial conditions (rather than using the first-period data) is only a little worse in hit rate than EWA and slightly worse in LL.

The bottom line of [Table 10.3](#), “pooled,” shows results when a single set of common parameters is estimated for all games (except for game-specific λ). If fEWA is capturing parameter differences across games effectively, it should predict especially accurately, compared to other models, when games are pooled. It does: when all games are pooled, fEWA 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 analyze all the data people sent us by the end of the year and report the results in a revised paper. Three samples were sent and we have analyzed one so far—experiments by [Kocher and Sutter \(2005\)](#) on p -beauty contest games played by individuals and groups. The KS results are reported in the bottom row of [Table 10.3](#). 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 fEWA , and reinforcement and QRE models fit worst. This is a small piece of evidence that the solid performance of fEWA (while worse than belief learning on these games) is not entirely due to overfitting on our original seven-game sample.

The table also shows results (in the column headed “Thinking+ fEWA ”) when the initial conditions are created by the CH model rather than from first-period data and combined with the fEWA learning model. Thinking plus fEWA are also a little more accurate than the belief and reinforcement models in five of seven games. The hit rate and LL suffer only a little compared to the fEWA with estimated parameters. When common parameters are estimated across games (the row labelled “pooled”), fixing initial conditions with the thinking steps model only lowers fit slightly.

Table 10.3 Out of sample accuracy of learning models (Ho et al. 2007)

Game	Thinking +fEWA			fEWA			EWA			Weighted Fict. Play			Reinf. with PV			QRE		
	%Hit	LL		%Hit	LL		%Hit	LL		%Hit	LL		%Hit	LL		%Hit	LL	
Cont'l divide	45	-483		47	-470		47	-460		25	-565		45	-557		5	-806	
Median action	71	-112		74	-104		79	-83		82	-95		74	-105		49	-285	
p-BC	8	-2119		8	-2119		6	-2042		7	-2051		6	-2504		4	-2497	
Price matching	43	-507		46	-445		43	-443		36	-465		41	-561		27	-720	
Mixed games	36	-1391		36	-1382		36	-1387		34	-1405		33	-1392		35	-1400	
Patent Race	64	-1936		65	-1897		65	-1878		53	-2279		65	-1864		40	-2914	
Pot Games	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	

Note: Sample sizes are 315, 160, 580, 160, 960, 1760, 739, 4674 (pooled), 80.

Note that all learning models have a low hit-rate in p -beauty contests (p -BC). This is so for three reasons. First, the p -BC has many strategies (101 altogether) and as a consequence it is much harder to predict behavior correctly in this game. Second, some subjects may be sophisticated and their behavior may not be captured by learning models (see [Section 10.5](#) for details). Finally, subjects may exhibit increasing sophistication over time (i.e., subjects increase their depth of reasoning) and adaptive learning models do not allow for this dynamic rule shift (see [Ho and Su, 2013a,b](#) for a dynamic model that explicitly accounts for such behavior).

10.4.4 Example: mixed strategy games

Some of the first games studied experimentally featured only mixed-strategy equilibria.²¹ These games usually have a purely competitive structure in which one person wins and another loses for each strategy combination (although they are not always zero-sum). This section summarizes the discussion in [Camerer \(2003, Chapter 3\)](#) and updates it.

The earliest results were considered to be discouraging rejections of mixed equilibrium. Most lab experiments do show non-independent over-alternation of strategies (i.e., subjects use fewer long runs than independence predicts). However, the conclusion that actual mixture frequencies were far from equilibrium proved to be too pessimistic. A survey of many experimental studies in [Camerer \(2003, Chapter 3\)](#) comparing actual *aggregate* choice frequencies with predicted frequencies shows they are highly correlated across all games ($r = 0.84$); the mean absolute deviation between predictions and data is 0.067. Fitting QRE instead of Nash mixtures improves the actual-predicted correlation a bit ($r = 0.92$) . . . and there is little more room for improvement because the correlation is so high!

Careful experiments by [Binmore et al. \(2001\)](#) and some field data on tennis and soccer (e.g., [Palacio-Huerta and Volij, 2008](#); [Walker and Wooders, 2001](#)) draw a similarly positive conclusion about overall accuracy of mixed equilibrium.²² So does an unusual example we describe next.

²¹ The earliest studies were conducted in the 1950s, shortly after many important ideas were consolidated and extended in [Von Neumann and Morgenstern's \(1944\)](#) landmark book. John Nash himself conducted some informal experiments during a famous summer at the RAND Corporation in Santa Monica, California. Nash was reportedly discouraged that subjects playing games repeatedly did not show behavior predicted by theory: "The experiment, which was conducted over a two-day period, was designed to test how well different theories of coalitions and bargaining held up when real people were making the decisions . . . For the designers of the experiment . . . the results merely cast doubt on the predictive power of game theory and undermined whatever confidence they still had in the subject" ([Nasar, 1998](#)).

²² The Palacios-Huerta and Volij study is unique because they are able to match data from actual play on the field from one group of players (in European teams) with laboratory behavior of some players from that group (although not matching the same players' field and lab data). Importantly, PHV also found that high school students as a group behaved less game theoretically than the soccer pros, except that high school students with substantial experience playing soccer were much closer to game-theoretic. However, a reanalysis by [Wooders \(2010\)](#) later showed a higher degree of statistical dependence than shown by PHV.

Table 10.4 Asymmetric mixed-strategy hide-seek games in Martin et al. (2013)

	L	R	Nash mse
L	X,0	0,2	$\frac{1}{2}$
R	0,2	Y,0	$\frac{1}{2}$
Nash	$\frac{Y}{X+Y}$	$\frac{X}{X+Y}$	

In the example study Martin et al. (2013) used three variants of asymmetric matching pennies (Table 10.4).

Subjects simultaneously chose L or R by pressing a touch screen display of a left or right box (like an iPad). The “Matcher” player wins either X or Y if both players match their choices on L or R, respectively. The “Mismatcher” player wins if they mismatch. Subjects participated in multiple sessions with 200 trials in each, switching matcher/mismatcher roles after each session. The games used (X,Y) matcher payoff pairs of (2,2), (3,1) and (4,1). NE makes the highly counterintuitive prediction that while the matcher payoffs change across the three games, the predicted $P_{\text{match}}(R) = 0.50$ does *not* change, but predicted $P_{\text{mismatch}}(R)$ does change, from 0.50 to 0.75 to 0.80. Data averaged over trials are shown in Figure 10.6 shows the averaged results from each of the three games, choosing over several hundred trials. The actual frequencies of $P_{\text{mismatch}}(R)$ The frequencies are remarkably close to the NE predictions. In five of the six comparisons, the absolute deviations are 0–0.03.

The experiments used six chimpanzees housed in the Kyoto PRI facility, playing in three mother–child pairs for apple cube payoffs delivered immediately. Also, we used a no-information protocol in which the chimpanzees were not told what the payoffs were; instead, they learned them from trial and error.

We can compare the chimpanzee behavior with two human groups, which separately played one of the three games, with $(X, Y) = (4, 1)$ (called “inspection game”), also using the low-information protocol. The chimpanzees are clearly closer to NE than the humans. The two human groups’ data are close together, despite large differences in the two groups (Japanese vs. African workers at a reserve in Boussou earning the equivalent of US \$134). The human subjects’ deviations are also similar in magnitude to many earlier experiments.

Why are the chimpanzees so close, and closer than humans? The apparent reason is that they learn better. Martin et al. estimated a weighted fictitious play model in which beliefs update according to $p(t+1)^* = p_t^* + \eta \delta_t^p$ where $\delta_t^p = P_t - p_t^*$ is the prediction error (the difference between observed play, 0 or 1, and belief). The mismatcher choice probability in each trial is $f(V^L - V^R)^{\text{Mismatcher}} = f([\pi(L, L)p^* + \pi(R, L)(1 - p^*)] - [\pi(L, R)p^* + \pi(R, R)(1 - p^*)])$ where $f(\cdot)$ is the logit function $f(V^a - V^b) = 1/(1 + e^{-\rho(V^a - V^b + \alpha)})$ (where α allows for bias toward L or R). When

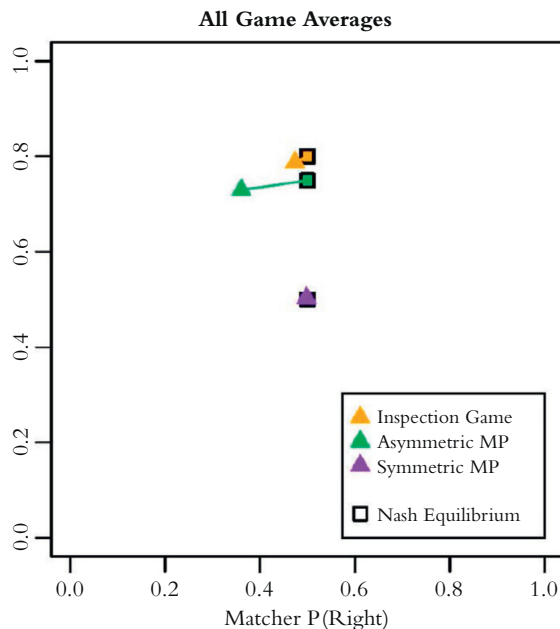


Figure 10.6 Actual relative frequencies of chimpanzee choices and predictions of Nash equilibrium, in three games.

fictitious play (or reinforcement learning²³) models are fit to the chimpanzee and human data, the chimpanzees show an enhanced sensitivity to past reward and opponent play (higher learning weight η , and to payoff structure (changing behavior between the matcher and mismatcher roles), compared to humans.

Martin et al. (2013) speculate that the chimpanzees are as good as the human subjects in mixing strategically (or perhaps even better in general), because chimpanzees may have retained the cognitive capacity to detect patterns in competitive games, and practice it more often in childhood, than humans do (Matsuzawa, 2007). Winning repeated competitions is important in the chimpanzees' natural ecology, in the form of hide-and-seek (predator-prey) games and winning physical competitions to establish status dominance. The learning model results suggest the chimpanzees learn better than the humans do. In general, the results strikingly supports the interpretation of (Binmore, 1999) of game theory as often “evolutionary”—that is, describing the result of an adaptive, perhaps biological process, rather than introspection and analysis.²⁴

²³ Note that EWA was not fit because in these types of games reinforcement, fictitious play and EWA are difficult to separate statistically

²⁴ This is a plausible explanation. However, there are other differences between the chimpanzee and human experiments. The most obvious difference is that chimpanzees receive instant gratification (e.g., they

10.4.5 Summary

Learning is clearly important for economics. Equilibrium theories are useful because they suggest a possible limit point of a learning process and permit comparative static analysis. But if learning is slow, or the time path of behavior selects one equilibrium out of many, a precise theory of equilibration is crucial for knowing which equilibrium will result, and how quickly.

The theory described in this section, fEWA, replaces the key parameters in the parametric EWA learning models with functions that change over time in response to experience. One function is a “change detector” ϕ which goes up (limited by one) when behavior by other players is stable, and dips down (limited by zero) when there is surprising new behavior by others. When ϕ dips down, the effects of old experience (summarized in attractions which average previous payoffs) is diminished by decaying the old attraction by a lot. The second “attention” function δ is one for strategies that yield equal or better than actual payoff and zero otherwise. This function ties sensitivity to foregone payoffs to attention, which is likely to be on strategies that give equal or better than actual payoff *ex post*. Self-tuning EWA is more parsimonious than most learning theories because it has only one free parameter—the response sensitivity λ .

We report out-of-sample prediction of data from several experimental games using fEWA, the parameterized EWA model, and QRE. Both QRE and self-tuning EWA have one free parameter, and EWA has five. We show that fEWA predicts slightly worse than parametric EWA in these games. Since fEWA generates functional parameter values for ϕ and δ , which vary sensibly across games, it predicts better than other QRE and parametric EWA when games are pooled and common parameters are estimated. fEWA therefore represents one solution to the problem of flexibly generating EWA-like parameters across games.

10.5. SOPHISTICATION AND TEACHING

The learning models discussed in the last 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 explicitly use information about *other* players’ payoffs (though subjects actually do²⁵); and adaptive models ignore the fact that when the same players are matched together repeatedly, their behavior is often different than then they are not rematched together, generally in the direction of greater efficiency (e.g., [Andreoni and Miller, 1993](#); [Clark and Sefton, 1999](#); [Van Huyck, et al., 1990](#)).

eat the apple cubes right away) and human subjects earn money that can only be traded for a delayed gratification. Such differences between the two experiments could also explain the observed differences in behaviors.

²⁵ [Partow and Schotter \(1993\)](#), [Mookerjee and Sopher \(1994\)](#), and [Cachon and Camerer \(1996\)](#).

In this section adaptive models are extended to include sophistication and strategic teaching in repeated games (see Camerer et al., 2002; and Stahl, 1999, for details). Sophisticated players believe that others are learning and anticipate how others will change in deciding what to do. In learning to shoot a moving target, for example, soldiers and fighter pilots learn to shoot *ahead*, toward where the target *will be*, rather than shoot at the current target. 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 be either 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 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 effects of information and matching that adaptive models miss.²⁶

10.5.1 Sophistication

Let’s begin with myopic sophistication (no teaching). The model assumes a population mixture in which a fraction α of players are sophisticated. To allow for possible overconfidence, sophisticated players think that a fraction $(1 - \alpha')$ of players are adaptive and the remaining fraction α' of players are sophisticated like themselves.²⁷ Sophisticated players use the fEWA 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

$$A_i^j(s, t) = \sum_{k=1}^{m-i} [\alpha' P_{-i}^k(s, t+1) + (1 - \alpha') \cdot P_{-i}^k(a, t+1)] \cdot \pi_i(s_i^j, s_{-i}^k) \quad [10.12]$$

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 QRE (and Nash equilibrium) are special cases in which all players are sophisticated and believe others are too ($\alpha = \alpha' = 1$).

An alternative model links steps of sophistication to the steps of thinking used in the first period. For example, define zero learning steps as using fEWA; one step is

²⁶ Sophistication may also potentially explain why players sometimes move in the *opposite* direction predicted by adaptive models (Rapoport et al., 1999), and why measured beliefs do not match up well with those predicted by adaptive belief learning models (Nyarko and Schotter, 2002).

²⁷ To truncate the belief hierarchy, the sophisticated players believe that the other sophisticated players, like themselves, believe there are α' sophisticates.

Table 10.5 Sophisticated and adaptive learning model estimates for the P -beauty contest game (Camerer et al., 2002) (standard errors are in parentheses)

	Inexperienced subjects		Experienced subjects	
	Sophisticated EWA	Adaptive EWA	Sophisticated EWA	Adaptive EWA
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

best-responding to zero-step learners; two steps is best-responding to choices of one-step sophisticates, and so forth (see Ho et al., 2013). 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 it reduces the entire thinking-learning-teaching model to only two parameters. In addition, this model allows players to become more or less sophisticated over time as they learn about others' thinking steps.

We estimate the sophisticated EWA model using data from p -beauty contests introduced above. Table 10.5 reports results and estimates of important parameters (with bootstrapped standard errors). 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% and their estimated perception $\hat{\alpha}'$ is zero, showing overconfidence (as in the thinking-steps estimates from the last section).²⁸

Experienced subjects are those who play a second 10-period game with a different p parameter (the multiple of the average which creates the target number). Among experienced subjects, the estimated proportion of sophisticates increases to $\hat{\alpha} = 77\%$. Their estimated perceptions increase too but are still overconfident ($\hat{\alpha}' = 41\%$). The estimates reflect “learning about learning”: Subjects who played one 10-period game come to realize an adaptive process is occurring; and most of them anticipate that others are learning when they play again.

²⁸ 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.

10.5.2 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 which will change what adaptive players do, in a way that benefits the sophisticated player in the long-run. Game theorists have showed 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 the paper 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 parsimoniously extended to explain behavior in more complex games which are, perhaps, of even greater economic interest than games with random matching.

Consider a finitely repeated trust game. 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 experiments are shown in [Table 10.6](#).

There are actually 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 that a given borrower is each type is commonly known). The honest (Y) types actually receive *more* money from repaying the loan, an experimenter’s way of inducing preferences like those of a person who has a social utility for being trustworthy (see [Camerer, 2003](#), Chapter 3 and references therein). 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.²⁹

Table 10.6 Payoffs in the Borrower-Lender trust game,
[Camerer and Weigelt \(1988\)](#)

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

²⁹ Note that player types need not be stage game types. They can also be repeated game types where players are preprogrammed to play specific repeated game strategies. We do not allow repeated game types in this Section.

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 actually play these sequential equilibria is important to investigate for two reasons. First, the equilibria impose consistency between optimal behavior 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 behavioral question (cf. [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 which are not confirmed in experiments (e.g., [Jung et al., 1994](#); [Neral and Ochs, 1992](#)).

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.³⁰ 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 10.7a](#) and b. 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 10.7a](#) shows that in early sequences lenders start by making loans in early periods (i.e., there is a low frequency of no-loan), but they rarely lend in periods 7–8. In later sequences they have learned to always lend in early periods and rarely lend in later periods. [Figure 10.7b](#) 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 teaching model is a boundedly rational model of reputation formation in which the lenders learn whether to lend or not. They do not update borrowers' types and do not anticipate borrowers' future behavior (as in equilibrium models); they just learn. In the teaching model, some proportion of borrowers are sophisticated and teach; the rest are adaptive and learn from experience but have no strategic foresight.

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

$$A_B^j(s, k, t) = \sum_{j'=\text{Loan}}^{\text{NoLoan}} P_L^{j'}(a, k, t+1) \cdot \pi_B(j, j') +$$

³⁰ 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.

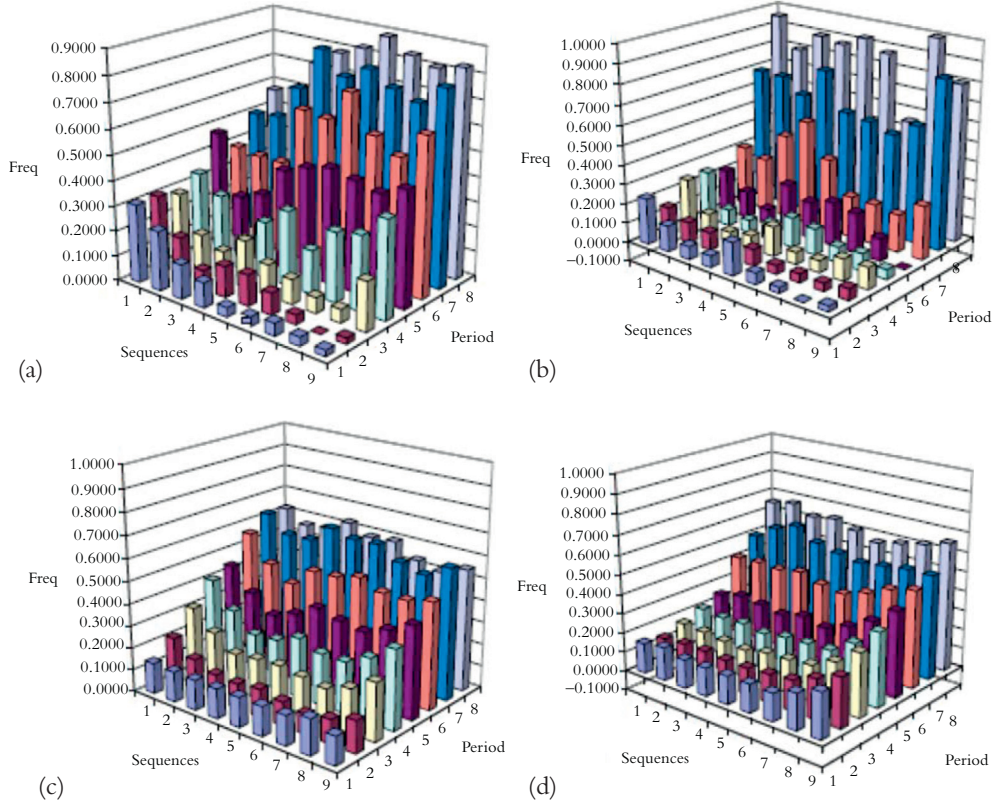


Figure 10.7 (a) Empirical frequency for no loan. (b) Empirical frequency for default conditional on loan. (c) Predicted frequency for no loan. (d) Predicted frequency for default conditional on loan.

$$\max_{J_{t+1}} \left\{ \sum_{v=t+2}^T \sum_{j'=\text{Loan}}^{\text{NoLoan}} \hat{P}_L^{j'}(a, k, v | j_{v-1} \in J_{t+1}) \cdot \pi_B(j_v \in J_{t+1}, j') \right\}$$

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 $j_{t+1} = j$.³¹ 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 periods v , which depends on what happened in

³¹ To economize in computing, we search only paths of future actions that always have default following repay because the reverse behavior (repay following default) generates a lower return.

the past (which the teacher anticipates).³² $P_B^j(s, k, t + 1)$ 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 fEWA with two twists. First, since lenders who play in later periods know what has happened earlier in a sequence, we assume that they learned from the experience they saw as if it had happened to them.³³ Second, a lender who is about to make a decision in period 5 of sequence 17, for example, has *two* relevant sources of experience to draw on—the behavior she saw in periods 1–4 in sequence 17, *and* the behavior she has seen in the period 5's of the previous sequences (1–16). Since *both* kinds of experience could influence her 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.³⁴ The parameter σ is a measure of the strength of “peripheral vision”—glancing back at the “future” period 5's from previous sequences to help guess what lies ahead.

Of course, it is well-known that repeated-game behavior can arise in finite-horizon games when there are a small number of “unusual” types (who act like the horizon is unlimited), which creates an incentive for rational players to behave as if the horizon is unlimited until near the end (e.g., [Kreps and Wilson, 1982](#)). But specifying why some types are irrational, and how many they 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; 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 [Ho et al. \(2007\)](#), we used parametric EWA to model behavior in trust games. That model allows two different sets of EWA parameters for lenders and borrowers. In this paper we use fEWA to model lenders and adaptive borrowers so the model has fewer

³² Formally, $\hat{P}_L^j(a, k, v|j_{v-1}) = \hat{P}_L^{\text{Loan}}(a, k, v - 1|j_{v-1}) \cdot P_L^j(a, k, v|(\text{Loan}, j_{v-1})) + \hat{P}_L^{\text{NoLoan}}(a, k, v - 1|j_{v-1}) \cdot P_L^j(a, k, v|(\text{NoLoan}, j_{v-1}))$.

³³ This is called “observational learning” (see [Duffy and Feltovich, 1999](#)). Without this assumption the model learns far slower than the lenders do so it is clear that they are learning from observing others.

³⁴ 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 the 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 [Ho et al. \(2007\)](#) for details.

Table 10.7 Model parameters and fit in repeated trust games

		Model	
		fEWA+ Teaching	Agent QRE
Statistic			
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	–
% of teachers	α	0.89	–
Homemade prior $p(\text{honest})$	θ	–	0.91

parameters.³⁵ Maximum likelihood estimation is used to estimate parameters on 70% of the sequences in each experimental session, then behavior in the holdout sample of 30% of the sequences is forecasted 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-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 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 design³⁶). (Standard equilibrium concepts are a special case of this model when λ 's are large and $\theta = 0$, and fit much worse than AQRE does.)

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 10.7 (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 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 of eight individual experimental

³⁵ 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 .

³⁶ See Camerer and Weigelt (1988), Palfrey and Rosenthal (1988), and McKelvey and Palfrey (1992).

sessions), predicting about 75% of the choices correctly. The AQRE fits reasonably well too (72% 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 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.

Figures 10.7c and 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 actually 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 as a boundedly rational alternative to type-based equilibrium models of reputation-formation.³⁷ 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 behavioral parameters (α and σ) appear to be near one, restricting attention to these values should make the model workable for doing theory.

10.5.3 Summary

In this section we introduced the possibility that players can be sophisticated (i.e., they believe others are learning). 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 a second time, as if 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 behavioral 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.

³⁷ One direction we are pursuing is to find designs or tests which distinguish the teaching and equilibrium updating approaches. The sharpest test is to compare behavior in games with types that are fixed across sequences with types that are independently “refreshed” in each period within a sequence. The teaching approach predicts similar behavior in these two designs but type-updating approaches predict that reputation formation dissolves when types are refreshed.

10.6. SOCIALITY

Some of the simplest economic games have proved to be useful as ways to measure sociality—departures from pure self-interest—and test specific models of social preferences. By far the most popular examples are prisoner's dilemma (or multiplayer "commons" dilemmas) and public goods contribution games (PGG). While these games are trivial from a strategic point of view, at first blush, explaining behavioral regularities has required fundamental developments and applications in game theory. Since the literature is large, we will discuss selective results from PGG, along with ultimatum and dictator games.

Before proceeding, however, it is crucial to clarify what sociality means in game theory. As every reader should know, game theory makes predictions about behavior conditional on a proper specification of utilities for outcomes. Rejecting money in an ultimatum game, for example, is definitely *not* a rejection of game theoretic rationality if the utility of having nothing is greater than the utility of having a small amount when a proposer has a much larger amount (due to fairness preferences, for example). However, since we typically do not have independent measurement of utilities for outcomes, in testing game theory predictions we are almost always testing the joint hypothesis of a particular utility specification—usually, selfishness—along with a theory of behavior given those specified utilities. The evidence discussed in this section shows systematic ways in which people do not appear to be selfish, and new ideas about what utility specifications can explain the data. In a sense, the goal of this approach is very "conservative"—the hope is that a reasonable utility specification will emerge so that data can be reasonably explained by standard analysis given that new type of utility.

It is also notable that the tendency of economists to be skeptical about human prosociality (e.g., Stigler, 1981; List quoted in Stanley, 2011) conflicts with the prevailing view of exceptional human prosociality, compared to other species, in biology and anthropology. Scientists in those fields have amassed much evidence that humans are more prosocial toward non-kin than all other species. For example, Boyd and Richerson wrote: "Humans cooperate on a larger scale than most other mammals.....The scale of human cooperation is an evolutionary puzzle" (Boyd and Richerson, 2009, p. 3281).

10.6.1 Public goods

In a typical linear public goods game experiment, four players are endowed with 20 tokens. A number c_i of tokens can be contributed to a public pool where they earn M , which is distributed equally. The rest of the tokens, $20 - c_i$, earn a private return of 1. The fraction $\frac{M}{4}$ is called the marginal per capita return (MCPR). Assumptions $1 < M < 4$ are imposed so that M is above the private return but the MCPR is below the private return. Player i 's payoff is $u_i(c_i, c_k) = 20 - c_i + \frac{M}{4} \sum_{j=1}^4 c_j$. The collective

payoff of $80 - (M - 1) \sum_{j=1}^4 c_j$ is maximized if $c_i = 20 \forall i$. Then each person earns $(20M * 4)/4 = 20M$. If everyone is selfish they contribute $c_i = 0$ and everyone earns 20.

Notice that under selfish preferences, keeping all tokens is a dominant strategy, so the predictions do not depend on the depth of strategic thinking. Therefore, the PGG is mostly interesting as a way to study social (nonselfish) preferences, and the interaction of such preferences with structural variables.

In most experiments, contributions begin with a bimodal distribution of many 100% and 0% contributions, averaging around 50% of the tokens. Given feedback about overall contributions, within 10 periods or so the average contribution erodes to about 10%. Contribution is substantially higher in a “partner” protocol in which the same people are matched each period, compared to a random-matching “stranger” protocol. If subjects are surprised and the entire experimental sequence is repeated, initial contributions jump back up near 50%. This “restart effect” (Andreoni, 1988) is clear evidence that subjects who contribute are not just confused but need experience to learn to give nothing, since they would give more after a restart.

What seems to be going on in the PGG is that some people are selfish, and give nothing no matter what. Another large fraction of people are “conditionally cooperative,” and give in anticipation that others will give. If they are too optimistic, or choose to punish the selfish free-riders, these conditional cooperators only recourse is to give less. These combined forces lead to a steady decline in contribution. The best evidence for conditional cooperation is an elegant experiment that asked subjects how much they would contribute, depending on how much other subjects did (and enforced actual contribution in this way; Fischbacher and Gächter, 2010). In this protocol, a large number of subjects choose to give only when others give.

10.6.2 Public goods with punishment

An interesting variant of PGG includes costly punishment after players contribute. The idea is that in natural settings players can often find out how much others contributed to a public good, and can then punish or reward other people based on their observed contributions, typically at a cost to themselves.

These punishments range from gossip, scolding, and public shaming (Guala, 2010)—now magnified by the internet—to paralegal enforcement (vigilantism, family or neighborhood members sticking up for each other), or organized legal enforcement through policing and taxation.

Yamagishi (1986) was the first to study punishment in the form of sanction. Since then, a paradigm introduced by Fehr and Gächter (2000) has become standard. In their method, after contributions are made a list of the contribution amounts is made public (without identifying identity numbers or other information about contributors, to avoid

repeated-game spillovers). Each player can then spend 1 token to subtract 3 tokens from one other participant, based only on the punished participants contribution.

The first wave of studies showed that punishment was effective in raising contributions close to the socially optimal step (although total efficiency suffers a bit since punishments reduce total payoff). When punishment is available, early punishment of free riders causes them to increase their contributions, and punishment is more rare in later trials as a result. Punishment works when some of those who contribute greatly to the public good also act as “social police,” using personal resources to goad low contributors to give more.

However later studies complicated this simple finding in several ways:

1. Some of the effect of punishment can be created by symbolic shaming, without any financial cost or penalty (Masclot et al., 2013).
2. Players have a choice with whom they wish to play the game (Page et al., 2005), what type of sanctions are allowed (Markussen et al., 2014), or whether or not to allow punishment in a community (Gurerk et al., 2006).
3. What happens if those who are punished can then punish the people who punish them? (Note that in natural settings, there is likely to be no way to allow informal punishment without retaliation against punishers.) One study showed that in this scenario, it is easy for efficiency-reducing punishment “feuds” to arise (Nikoforakis and Engelmann, 2011).
4. Herrmann et al. (2008) and Gächter et al. (2010) found an empirical continuum of punishment in an amazing study across 16 societies, sampling a wider variety of cultures than previously studied. In countries with English-speaking, Confucian, or Protestant histories, those who punish a lot are typically high contributors punishing low contributors (punishment is “prosocial” in intent). However, in several Arabic, Mediterranean, and former Communist block countries they sampled, there were also high rates of “antisocial” punishment (i.e., low contributors often punish high contributors)! When antisocial punishment is common, giving rates do not increase across trials, and efficiency is reduced because punishments are costly for both parties involved.

10.6.3 Negative reciprocity: ultimatums

In a typical ultimatum game, a proposer offers a specific division of a known amount of money to a responder. The responder can accept it, and the money amounts are paid to both sides, or reject it, and both sides get nothing.

This simple game was first studied experimentally by Guth et al. (1982) and has been replicated hundreds of times since then (including offers of water to very thirsty people; Wright et al., 2012). If players are self-interested, then the subgame perfect equilibrium is to offer zero (or the smallest divisible amount) and for responders to accept. This joint hypothesis about strategic thinking is typically rejected because average offers are

around 37%, and offers below 20% or so are often rejected. (The overall rejection rate was 16%, in a survey of 75 studies by Oosterbeek et al., 2004.) Since the responders rejection choice does not involve any strategic thinking at all, these stylized facts are clear evidence that many players deciding whether to reject or accept offers are not purely self-interested.

One explanation that emerged prominently early on is that rejections are simple monetary mistakes which selfish players learn to avoid over time (e.g., Erev and Roth, 1998). This learning explanation is clearly incomplete: It does not specify the time scale over which learning takes place, and cannot explain why fewer rejection “mistakes” are made when unequal offers are generated exogenously by a computer algorithm or when rejections do not hurt the proposer. Evidence for learning across repeated experimental trials is also not strong (e.g., Cooper and Stockman, 2002).

The leading current explanation for ultimatum rejections is “negative reciprocity”: Some players trade off money with a desire to deprive a player who has treated them unfairly from earning money. Summarized from strongest to weakest, evidence for a negative reciprocity explanation comes from the facts that: players reject offers less frequently if an unequal offer is created by a computerized agent (Blount, 1995; Sanfey et al., 2003); rejections are rare if rejecting the offer does not deprive the proposer of her share (Bolton and Zwick, 1995); low offers and rejections are associated with brain activity in insula, a bodily discomfort region (Sanfey et al., 2003), and with wrinkling facial muscles associated with felt anger and disgust (Chapman et al., 2009) or exogenously induced disgust (Moretti and di Pellegrino, 2010); and rejections are associated with self-reported anger (Pillutla and Murnighan, 1996).

Many studies have changed the amount of money being divided, over substantial ranges including days or weeks of wage equivalents (beginning with Cameron, 1995). Offers generally go down as the stakes go up (Oosterbeek et al., 2004). As stakes go up, the absolute amounts players will reject increase, but the percentages they reject decrease. Note that this pattern does *not* imply responders are becoming “more rational” as stakes increase, because the money amounts they will reject increase in magnitude. In fact, such a pattern is predicted to arise if responders care both about their own earnings and about their relative share (or if equality is an inferior good).

Several formalizations of social preference have been developed to explain rejected offers. The simplest of these propose utility functions that include aversion to unequal outcome differences (Fehr and Schmidt, 1999) or unequal shares (Bolton and Ockenfels, 2000). However, these theories cannot explain acceptance of low offers generated exogenously (which create equal inequity). Therefore, other theories model ultimatums as psychological games in which beliefs enter the utility function (Rabin, 1993).

An interesting empirical strategic question is whether proposers have accurate beliefs about how much responders are willing to accept, and make payoff-maximizing offers given those beliefs. Early findings in four countries from Roth et al. (1991) support this

“rational proposer” hypothesis, but without formal statistical testing or adjustment for risk-aversion. This hypothesis is an attractive one for game theory because it suggests players have equilibrium beliefs and choose best responses (or noisy quantal responses) given their beliefs. If so, after adjusting for social preferences or negative reciprocity, equilibrium analysis is an accurate description of offers and rejection rates.

However, several later studies found domains in which proposers appear to offer too much or too little, as evidenced by apparent gaps between actual offers and profit-maximizing offers given accurate beliefs.

For example, many small-scale societies show generous near-equal offers along with extremely low rejection rates at all offer steps, suggesting that proposer beliefs are pessimistic (Henrich et al., 2001). Offers in a study in the Khasi-speaking part of India (Andersen et al., 2011) are half as large in most samples (12–23%) but rejection rates are higher than in any country ever studied, around (34%). Either the Khasi proposers’ beliefs about acceptance are too optimistic, or they are extremely risk-seeking.

10.6.4 Impure altruism and social image: dictator games

A dictator game is a simple decision about how to allocate unearned money between the dictator and some recipient(s). It is an ultimatum game with no responder rejection. Dictator games were first used to see whether ultimatum offers are altruistic (if so, dictator and ultimatum offers will be the same) or are selfishly chosen to avoid rejection (if so, case dictator offers will be lower). Dictator offers are clearly lower.

However, dictator game allocations are sensitive to many details of procedure and (likely) personality. In psychological terms, the dictator game is a “weak situation” because norms of reasonable sharing can vary with context, may not be widely shared, and compete with robust self-interest.

In initial dictator game experiments some subjects give nothing but a small majority do allocate some money to recipients; the average allocation is typically 10–25%. Subjects clearly give much less when the money to be allocated was earned by the dictator rather than unearned and randomly allocated by the experimenter (Camerer and Thaler, 1995, p. 216; cf. Hoffman and Spitzer, 1982). Therefore, it should be clear that experimental dictator giving in lab settings is likely to be much more generous than natural charitable giving out of wealth. Allocations are also sensitive to the range of what can be given or taken (Bardsley, 2008; List, 2007), and to endogenous formation of norms about what others give (Krupka and Weber, 2009).

In the canonical dictator game, the total allocation amount is fixed, so there is no tradeoff between giving more and giving more equally. However, if the possible total allocations vary, then it is useful to know how much people will sacrifice efficiency (choosing a higher total allocation) for equity (choosing a more equal allocation). Several experiments explored this tradeoff (Charness and Rabin, 2002; Engelmann and Strobel, 2006). There is substantial evidence for efficiency-enhancing motivations, which appears

in some studies to depend on the subjects' academic background and training (Fisman et al., 2007).

The last several years have seen a correction to the language used to describe dictator game allocations imply. Some early discussions of these results called dictator giving "altruistic," because of the contrast to strategic offers in ultimatum games (which are not generous, per se, but instead meant to avoid rejections). Of course, the term "altruism" means something specific in public finance: Pure altruism is a desire to make another person better-off on that person's own terms. Impure altruism means having a private motive to help another person, such as a "warm glow" from giving.) In addition, [Camerer and Thaler \(1995\)](#) described dictator allocations as reflecting "manners," meaning adherence to an accepted social norm of giving.

The idea of "manners" was made precise in formal theories of "social image." In these theories, people give in dictator games because they have a utility for believing that others think they are a generous type (i.e., have good manners) or to avoid disappointing the expectations that they think other people have (guilt-aversion; [Hoffman et al., 1996](#)). Andreoni and Bernheim adapted a strategic theory of rational conformity to this setting. It predicts that if there is a P chance a dictator allocation will be stopped, so the recipient gets zero, then dictators will be more inclined to give 0 if P is higher. (The intuition is that if recipients are likely to get nothing anyway, the dictators won't feel as unfairly judged if they give zero also.) Indeed, they find such an effect in their experiment. [Berman and Small \(2012\)](#) also find that when the selfish choice is imposed exogeneously, subjects are happier. [Tonin and Vlassopoulos \(2013\)](#) found that a third of subjects who initially gave a positive amount change their minds when asked if they want to keep all the money. [Malmendier et al. \(2012\)](#) found that many subjects would rather opt out of a \$10 ultimatum game, to get \$9, provided the potential recipient subject would not know about their choice.³⁸

These studies indicate that in simple dictator allocations, a major motivation for dictator giving is wanting to *appear* generous, in order to believe that a recipient subject think the dictator is being generous. (In formal terms, people have a preference for an iterated belief that another person believes their type is generous.) Disrupting any element of this complicated chain of perception and cognition can easily push giving toward zero. Interesting biological evidence also supports the social image view: Contributing when others are watching activates reward areas ([Izuma et al. 2008](#)), and autistic adults do not adjust giving when being watched (which control subjects do).

It is always useful to ask how well findings from lab experiments correlate with field data when relevant features such as incentives, subject pools, choice sets etc. are closely matched (Camerer, in press). For prosociality, lab-field correlation is over-

³⁸ There are also studies showing that dictators may take away other player's money in order to equalize payoffs between them and others ([Zizzo, 2003](#); [List, 2007](#)).

whelmingly positive and often very impressive in magnitude (comparable to other kinds of psychometric reliability). For example, Brazilians who catch fish and shrimp exhibit lab prosociality and patience which is associated with more cooperative behavior in their fishing (Fehr and Leibbrandt, 2011). Boats of Japanese fishermen who disapprove of how much others keep selfishly tend to catch more collectively (Carpenter and Seki, 2011). Students who are generous in the (lab) dictator game are also more inclined to return a (field) misdirected letter containing money (Franzen and Pointner, 2013).

10.6.5 Summary

There is now massive evidence of various kinds of prosocial behavior. The challenge is to distill this evidence into a minimal set of workable, psychologically plausible models that can be used in public finance, game theory, political science and other applications. Leading theories assume simple kinds of adjustment to outcome utility based on inequity of payoff differences or shares, reciprocity, or social image. All theories are backed by various kinds of data. Notably, the data comparing lab measures of prosociality and field measures (holding as many lab-field factors constant as possible) shows a substantial correspondence of behavior in the lab and field.

10.7. CONCLUSION

This chapter presented experimental evidence motivating behavioral models based on limited rationality, stochastic choice, learning and teaching, and prosocial preferences. These models are intended to explain behavior in many different games with a single general specification (sometimes up to one or more free parameters).

Please note that not all examples illustrate how deviations from conventional equilibrium are explained by behavioral game theory. In some cases, the fact that equilibrium concepts predict surprisingly well can be explained by behavioral models. One example is entry games, which approximately equilibrate with no learning at all because, we suggest, higher-step thinkers fill in deviations from equilibrium that result from lower-step thinkers. Another example is how responsive learning by highly trained chimpanzees in hide-and-seek games (perhaps a conserved cognitive capacity due to its important for fitness in their competitive societies) leads to near-perfect frequencies of Nash play.

While these behavioral models are actively researched and applied, we are struck by how little attention is currently paid to them in typical game theory courses and textbooks. If a student asks, “What happens when people actually play these games?” equilibrium analysis should be part of any explanation, but it will typically not provide the best currently available answer to the student’s question. Many of the behavioral ideas are also easy to teach and easily grasped by students—partly because the intuitions

underlying the models correspond to students' intuitions about they would do or expect. And there are plenty of experimental data to contrast equilibrium predictions with intuitions (or to show they may match surprisingly well).

Many interesting behavioral and experimental topics were excluded in this chapter to save space. One set of models are boundedly rational solutions concepts which predict stationary strategy distributions (typically using one free parameter.) These include action sampling, payoff sampling, and impulse-balance theory (Chmura et al., 2008) Note that those solution concepts were not derived a priori from any specific evidence about cognitive limits (and have not been tested by eyetracking or fMRI data). There is also no robust evidence that these concepts improve on QRE when errors in Chmura et al. (2008) were corrected (see Brunner et al., 2011).

We also acknowledge cursed equilibrium (CE, Eyster and Rabin, 2008) and analogy-based expectational equilibrium (ABEE, Jehiel, 2005). There are very few new data testing these theories and, despite their potential, their current scope and applicability are limited. CE is specifically designed only to explain deviations in BNE with private information, so it coincides with Nash equilibrium in games with complete information. As a result, it is not a serious candidate for a general theory of deviation from Nash equilibrium across all types of games. ABEE predictions depend on how analogies are created, which requires auxiliary assumptions.

Frontiers of behavioral and experimental game theory are pushing far out from conventional lab choices, to tests with field data. They are also pushing far in and down, to show cognitive and neural mechanisms. We believe it is possible to have theories, such as CH, which makes specific predictions about mental processes *and* can be abstracted upward to make predictions about price, quantity and quality in markets.

REFERENCES

- Alaoui, L., Penta, A., 2013. Depth of reasoning and incentives. Econ. Wisc working paper.
- Andersen, S., Erta, S., Gneezy, U., Hoffman, M., List, J.A., 2011. Stakes matter in ultimatum games. *Am. Econ. Rev.* 101, 3427–3439.
- Andreoni, J., 1988. Why free ride? Strategies and learning in public goods experiments. *J. Public Econ.* 37, 291–304.
- Andreoni, J., Miller, J., 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. *Econ. J.* 103, 570–585.
- Arthur, B., 1991. Designing economic agents that act like human agents: a behavioral approach to bounded rationality. *Am. Econ. Rev.* 81 (2), 353–359.
- Bardsley, N., 2008. Dictator game giving: altruism or artifact. *Exp. Econ.* 11, 122–133.
- Berman, J.Z., Small, D.A., 2012. Self-interest without selfishness: the hedonic benefit of imposed self-interest. *Psychol. Sci.* 23 (10), 1193–1199.
- Binmore, K., 1999. Why experiment in economics? *The Econ. J.* 109 (453), 16–24.
- Binmore, K., Swierzbinski, J., Proulx, C., 2001. Does maximin work? An experimental study. *Econ. J.* 111, 445–464.

- Blount, S., 1995. When social outcomes aren't fair: the effect of causal attributions on preferences. *Org. Behav. Human Decis. Processes* 63, 131–144.
- Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90 (1), 166–193.
- Bolton, G.E., Zwick, R., 1995. Anonymity versus punishment in ultimatum bargaining. *Games Econ. Behav.* 10 (1), 95–121.
- Boyd, R., Richerson, P.J., 2009. Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. B* 364, 3281–3288.
- Brown, G., 1951. Iterative Solution of Games by Fictitious Play. In *Activity Analysis of Production and Allocation*. John Wiley & Sons, New York.
- Brunner, C., Camerer, C., Goeree, J., 2011. Stationary concepts for experimental 2×2 games: comment. *Am. Econ. Rev.* 101, 1029–1040.
- Bulow, J., Klemperer, P., 2002. Prices and the winner's curse. *RAND J. Econ.* 33 (1), 1–21.
- Bush, R., Mosteller, F., 1955. *Stochastic Models for Learning*. John Wiley & Sons, New York.
- Cabrales, A., Garcia-Fontes, W., 2000. Estimating learning models with experimental data. University of Pompeu Febrà Working Paper.
- Cachon, G.P., Camerer, C.F., 1996. Loss-avoidance and forward induction in experimental coordination games. *Q. J. Econ.* 111 (1), 165–194.
- Camerer, C.F., 1987. Do biases in probability judgment matter in markets? *Experimental Evidence*. *Am. Econ. Rev.* 77, 981–997.
- Camerer, C.F., 1995. Individual decision making. In: *Handbook of Experimental Economics*. Princeton University Press, Princeton, pp. 587–703.
- Camerer, C.F., 2003. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press, Princeton.
- Camerer, C.F., The promise and success of lab-field generalizability in experimental economics: a critical reply to Levitt and List. In: Frechette, G., Schotter, A. (Eds.), *The Methods of Modern Experimental Economics*. Oxford University Press, in press.
- Camerer, C.F., Fehr, E., 2006. When does “economic Man” Dominate Social Behavior?. *Science* 311 (5757), 47–52.
- Camerer, C.F., Weigelt, K., 1988. Experimental tests of a sequential equilibrium reputation model. *Econometrica* 56, 1–36.
- Camerer, C.F., Weigelt, K., 1993. Convergence in experimental double auctions for stochastically lived assets. In: Friedman, D., Rust, J. (Eds.), *The Double Auction Market: Theories, Institutions and Experimental Evaluations*. Addison-Wesley, Redwood City, CA, pp. 355–396.
- Camerer, C.F., Thaler, R.H., 1995. Anomalies: ultimatums, dictators and manners. *J. Econ. Perspect.* 9 (2), 209–219.
- Camerer, C.F., Ho, T.-H., 1998. EWA learning in normal-form games: probability rules, heterogeneity and time variation. *J. Math. Psychol.* 42, 305–326.
- Camerer, C.F., Ho, T.-H., 1999. Experience-weighted attraction learning in normal-form Games. *Econometrica* 67, 827–874.
- Camerer, C.F., Ho, T.-H., Chong, J.K., 2002. Sophisticated EWA learning and strategic teaching in repeated games. *J. Econ. Theory* 104 (1), 137–188.
- Camerer, C.F., Hsia, D., Ho, T.-H., 2002. EWA learning in bilateral call markets. In: Rapoport, A., Zwick, R. (Eds.), *Experimental Business Research*, pp. 255–284.
- Camerer, C., Ho, T.-H., Chong, J.K., 2004. Behavioural game theory: thinking, learning and teaching. In: *Advances in Understanding Strategic Behaviour: Game Theory, Experiments, and Bounded Rationality: Essays in Honor of Werner Güth*. Palgrave Macmillan, New York, pp. 120–180.
- Capra, M., 1999. Noisy expectation formation in one-shot games. Unpublished dissertation, University of Virginia.
- Capra, M., Goeree, J., Gomez, R., Holt, C., 1999. Anomalous behavior in a traveler's dilemma. *Am. Econ. Rev.* 89, 678–690.
- Carpenter, J., Seki, E., 2011. Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama bay. *Econ. Inquiry* 49 (2), 612–630.

- Carrillo, J.D., Palfrey, T.R., 2011. No trade. *Games Econ. Behav.* 71 (1), 66–87.
- Chapman, H.A., Kim, D.A., Susskind, J.M., Anderson, A.K., 2009. In bad taste: evidence for the oral origins for moral disgust. *Science*, 27, 1222–1226.
- Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. *Q. J. Econ.* 117 (3), 817–869.
- Cheung, Y.-W., Friedman, D., 1997. Individual learning in normal form games: some laboratory results. *Games Econ. Behav.* 19, 46–76.
- Chmura, T., Sebastian, G., Selten, R., 2007. Stationary concepts for experimental 2x2-games. *Am. Econ. Rev.* 98(3), 938–66.
- Chong, J.K., Camerer, C., Ho, T.H., 2004. Cognitive Hierarchy: A Limited Thinking Theory of Games, Experimental Business Research, Vol. II. In: Zwick, R., Rapoport, A. (Eds.), Kluwer Academic Publishers.
- Clark, K., Sefton, M., 1999. Matching protocols in experimental games. University of Manchester working paper.
- Cooper, D.J., Stockman, C.K., 2002. Learning to punish: experimental evidence from a sequential step-level public goods game. *Exp. Econ.* 5 (1), 39–51.
- Costa-Gomes, M., Crawford, V., Broseta, B., 2001. Cognition and behavior in normal-form games: an experimental study. *Econometrica* 69, 1193–1235.
- Crawford, V., 1997. Theory and experiment in the analysis of strategic interactions. In: Kreps, D., Wallis, K. (Eds.), *Advances in Economics and Econometrics: Theory and Applications. Seventh World Congress*, Vol. I. Cambridge University Press, Cambridge.
- Crawford, V., Costa-Gomes, M.A., Iriberry, N., 2013. Structural models of nonequilibrium strategic thinking: theory, evidence, and applications. *J. Econ. Lit.* 51, 5–62.
- Cross, J., 1983. *A theory of adaptive learning economic behavior*. Cambridge University Press, New York.
- Devetag, G., Warglien, M., 2003. Games and phone numbers: Do short-term memory bounds affect strategic behavior? *J. Econ. Psych.* 24, 189–202.
- Duffy, J., Feltovich, N., 1999. Does observation of others affect learning in strategic environments? An experimental study. *Int. J. Game Theory* 28, 131–152.
- Engelmann, D., Strobel, M., 2006. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: reply. *Am. Econ. Rev.* 96 (5), 1918–1923.
- Erev, I., Roth, A., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed-strategy equilibria. *Am. Econ. Rev.* 88, 848–881.
- Erev, I., Bereby-Meyer, Y., 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. *J. Econ. Behav. Org.* 39, 111–128.
- Erev, I., Roth, A.E., Slonim, R.L., Barron, G., 2002. Predictive value and the usefulness of game theoretic models. *Int. J. Forecast.* 18 (3), 359–368.
- Eyster, E., Rabin, M., 2005. Cursed equilibrium. *Econometrica* 73, 1623–1672.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114 (3), 817–868.
- Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* 90 (4), 980–994.
- Fehr, E., Leibbrandt, A., 2011. A field study on cooperativeness and impatience in the tragedy of the commons. *J. Public Econ.* 95 (9–10), 1144–1155.
- Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* 100 (1), 541–556.
- Fishman, R., Kariv, S., Markovits, D., 2007. Individual preferences for giving. *Am. Econ. Rev.* 97, 1858–1876.
- Franzen, A., Pointner, S., 2013. The external validity of giving in the dictator game: a field experiment using the misdirected letter technique. *Exp. Econ.* 16, 155–169.
- Fudenberg, D., Levine, D., 1989. Reputation and equilibrium selection in games with a patient player. *Econometrica* 57, 759–778.
- Fudenberg, D., Levine, D., 1998. *The Theory of Learning in Games*. MIT Press, Boston.

- Gächter, S., Herrmann, B., Thoni, C., 2010. Culture and cooperation. *Phil. Trans. R. Soc. B* 365 (1553), 2651–2661.
- Georgiadis, S., Healy, P.J., Weber, R., 2013. On the persistence of strategic sophistication. University of London Working Paper.
- Gill, D., Prowse, V., 2012. Cognitive ability and learning to play equilibrium: a level-k analysis. Munich Personal RePEc Archive, Paper No. 38317, posted 23 April, 2012. <http://mpra.ub.uni-muenchen.de/38317/>
- Goeree, J.K., Holt, C.A., 1999. A theory of noisy introspection. University of Virginia Department of Economics.
- Goeree, J.K., Holt, C.A., 2001. Ten little treasures of game theory, and ten contradictions. *Am. Econ. Rev.* 91 (5), 1402–1422.
- Goeree, J.K., Holt, C.A., Palfrey, T.R., 2005. Regular quantal response equilibrium. *Exp. Econ.* 8, 347–367.
- Guala, F., 2010. Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. University of Milan Department of Economics, Business and Statistics Working Paper No. 2010–23.
- Gürerk, O., Irlenbusch, B., Rockenbach, B., 2006. The competitive advantage of sanctioning institutions. *Science* 312 (5770), 108–111.
- Guth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Org.* 3 (4), 367–388.
- Harley, C., 1981. Learning the evolutionary stable strategies. *J. Theor. Biol.* 89, 611–633.
- Hart, S., Mas-Colell, A., 2001. A reinforcement procedure leading to correlated equilibrium. In: Debreu G., Neufeld W., Trockel (Eds.), *Economics Essay*, Springer Berlin Heidelberg, 181–200.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C.F., Fehr, E., 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* 91 (2), 73–78.
- Herrmann, B., Thoni, C., Gächter, S., 2008. Antisocial punishment across societies. *Science* 319 (5868), 1362–1367.
- Ho, T.-H., Weigelt, K., 1996. Task complexity, equilibrium selection, and learning: an experimental study. *Manage. Sci.* 42, 659–679.
- Ho, T.-H., Su, X., 2013. A dynamic level-k model in sequential games. *Manage. Sci.* 59, 452–469.
- Ho, T.-H., Camerer, C.F., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental ‘p-beauty contests’. *Am. Econ. Rev.* 88, 947–969.
- Ho, T.-H., Camerer, C.F., Chong, J.K., 2007. Self-tuning experience-weighted attraction learning in games. *J. Econ. Theory* 133, 177–198.
- Ho, T.-H., Wang, X., Camerer, C.F., 2008. Individual differences in the EWA learning with partial payoff information. *Econ. J.* 118, 37–59.
- Ho, T.-H., Park, S.-E., Su, X., 2013. Level-r model with adaptive and sophisticated learning. University of California, Berkeley Working Paper.
- Hoffman, E., Spitzer, M.L., 1982. The coase theorem: some experimental tests. *J. Law Econ.* 25 (1), 73–98.
- Hoffman, E., McCabe, K., Smith, V.L., 1996. Social distance and other regarding behavior in dictator games. *Am. Econ. Rev.* 86, 653–660.
- Hopkins, E., 2002. Two competing models of how people learn in games. *Econometrica* 70 (6), 2141–2166.
- Hyndman, R.J., Booth, H., Yasmeen, F., 2013. Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50 (1), 261–283.
- Itti, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40 (10–12), 1489–1506.
- Ivanov, A., Levin, D., Niederle, M., 2010. Can relaxation of beliefs rationalize the winner’s curse? An experimental study. *Econometrica* 78 (4), 1435–1452.
- Izuma, K., Saito, D.N., Sadato, N., 2008. Processing of social and monetary rewards in the human striatum. *Neuron* 58 (2), 284–294.
- Jehiel, P., 2005. Analogy-based expectation equilibrium. *J. Econ. Theory* 123, 81–104.
- Jung, Y.J., Kagel, J.H., Levin, D., 1994. On the existence of predatory pricing: an experimental study of reputation and entry deterrence in the chain-store game. *RAND J. Econ.* 25, 72–93.
- Kahneman, D., 1988. Experimental economics: a psychological perspective. In: Tietz, R., Albers, W., Selten, R., (Eds.), *Bounded Rational Behavior in Experimental Games and Markets*. Springer-Verlag, New York. pp. 11–18.

- Keynes, J.M., 1936. *The General Theory Of Employment, Interest And Money*. Macmillan Cambridge University Press, for Royal Economic Society.
- Kneeland, T., 2013. Rationality and consistent beliefs: theory and experimental evidence. University of British Columbia Working Paper.
- Kocher, M.G., Sutter, M., 2005. When the 'decision maker' matters: individual versus team behavior in experimental 'beauty-contest' games. *Econ. J.* 115, 200–223.
- Kreps, D., Wilson, R., 1982. Reputation and imperfect information. *J. Econ. Theory* 27, 253–279.
- Krupka, E., Weber, R.A., 2009. The focusing and informational effects of norms on pro-social behavior. *J. Econ. Psychol.* 30 (3), 307–320.
- List, J.A., 2007. On the interpretation of giving in dictator games. *J. Polit. Econ.* 115 (3), 482–493.
- Lucas, R.G., 1986. Adaptive behavior and economic theory. *J. Bus.* 59, S401–S426.
- Malmendier, U., Lazear, E., Weber, R., 2012. Sorting in experiments with application to social preferences. *Am. Econ. J. Appl. Econ.* 4 (1), 136–163.
- Markussen, T., Putterman, L., Tyran, J.-R., 2014. Self-organization for collective action: an experimental study of voting on sanction regimes. *Rev. Econ. Stud.* 81(1), 301–324.
- Martin, C., Bossaerts, P., Matsuzawa, T., Camerer, C., 2014. Experienced chimpanzees play according to game theory. *Nature Scientific Reports*, 4(5182).
- Masclét, D., Noussair, C.N., Villeval, M., 2013. Threat and punishment in public goods experiments. *Econ. Inquiry* 51 (2), 1421–1441.
- Matsuzawa, T., 2007. Comparative cognitive development. *Dev. Sci.* 10, 97–103.
- McAllister, P.H., 1991. Adaptive approaches to stochastic programming. *Ann. Operat. Res.* 30, 45–62.
- McKelvey, R.D., Palfrey, T.R., 1992. An experimental study of the centipede game. *Econometrica* 60, 803–836.
- McKelvey, R.D., Palfrey, T.R., 1995. Quantal response equilibria for normal-form games. *Games Econ. Behav.* 7, 6–38.
- McKelvey, R.D., Palfrey, T.R., 1998. Quantal response equilibria for extensive-form games. *Exp. Econ.* 1, 9–41.
- Moinas, S., Pouget, S., 2013. The bubble game: an experimental study. *Econometrica* 81 (4), 1507–1539.
- Mookerjee, D., Sopher, B., 1994. Learning behavior in an experimental matching pennies game. *Games Econ. Behav.* 7, 62–91.
- Mookerjee, D., Sopher, B., 1997. Learning and decision costs in experimental constant-sum games. *Games Econ. Behav.* 19, 97–132.
- Moretti, L., Di Pellegrino, G., 2010. Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion* 10 (2), 169–180.
- Myerson, R.B., 1998. Population uncertainty and poisson games. *Int. J. Game Theory* 27, 375–392.
- Nagel, R., 1995. Experimental results on interactive competitive guessing. *Am. Econ. Rev.* 85, 1313–1326.
- Nagel, R., Bosch-Domenech, A., Satorra, A., Garcia-Montalvo, J., 1999. One, two, (three), infinity: newspaper and lab beauty-contest experiments. Universitat Pompeu Fabra.
- Nasar, S., 1998. *A beautiful mind: a biography of John Forbes Nash Jr.* Simon and Schuster, New York.
- Neral, J., Ochs, J., 1992. The sequential equilibrium theory of reputation building: a further test. *Econometrica* 60, 1151–1169.
- Nikiforakis, N., Engelmann, D., 2011. Altruistic punishment and the threat of feuds. *J. Econ. Behav. Org.* 78 (3), 319–332.
- Nunnari, S., Camerer, C.F., Palfrey, T., 2012. Quantal response and nonequilibrium beliefs explain overbidding in maximum-value auctions. Caltech Working Paper.
- Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. *Econometrica* 70 (3), 971–1005.
- Ochs, J., 1999. Entry in experimental market games. In: Burescu, D., Erev, I., Zwick, R. (Eds.), *Games and Human Behavior: Essays in Honor of Amnon Rapoport*. Lawrence Erlbaum Assoc. Inc., New Jersey.
- Oosterbeek, H., Sloof, R., van de Kuilen, G., 2004. Differences in ultimatum game experiments: evidence from a meta-analysis. *Exp. Econ.* 7 (2), 171–188.
- Ostling, R., Wang, J.T.-y., Chou, E.Y., Camerer, C.F., 2011. Testing game theory in the field: Swedish LUPU lottery games. *Am. Econ. J. Microecon.* 3 (3), 1–33.

- Palacios-Huerta, I., Volij, O., 2008. Experientia docet: professionals play minimax in laboratory experiments. *Econometrica* 76 (1), 71–115.
- Palfrey, T.R., Rosenthal, H., 1988. Private incentives in social dilemmas: the effects of incomplete information and altruism. *J. Public Econ.* 35, 309–332.
- Page, T., Putterman, L., Unel, B., 2005. Voluntary association in public goods experiments: reciprocity, mimicry, and efficiency. *Econ. J.* 115, 1032–1053.
- Partow, J., Schotter, A., 1993. Does game theory predict well for the wrong reasons? An experimental investigation. C.V. Starr Center for Applied Economics Working Paper, New York University, pp. 93–46.
- Pilludla, M.M., Murnighan, J.K., 1996. Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Org. Behav. Human Decis. Processes* 68 (3), 208–224.
- Rabin, M., 1993. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83 (5), 1281–1302.
- Rapoport, A., Amaldoss, W., 2000. Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge. *J. Econ. Behav. Org.* 42, 483–521.
- Rapoport, A., Lo, A., K.-C., Zwick, R., 1999. Choice of prizes allocated by multiple lotteries with endogenously determined probabilities. University of Arizona, Department of Management and Policy Working Paper.
- Rogers, B.W., Palfrey, T.R., Camerer, C.F., 2009. Heterogeneous quantal response equilibrium and cognitive hierarchies. *J. Econ. Theory* 144 (4), 1440–1467.
- Rosenthal, R.W., 1981. Games of perfect information, predatory pricing and the chain-store paradox. *J. Econ. Theory* 25, 92–100.
- Roth, A., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games Econ. Behav.* 8, 164–212.
- Roth, A., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. *Am. Econ. Rev.* 81 (5), 1068–1095.
- Salmon, T., 1999. Evidence for 'learning to learn' behavior in normal-form games. Caltech Working Paper.
- Salmon, T., 2001. An evaluation of econometric models of adaptive learning. *Econometrica* 69, 1597–1628.
- Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758.
- Seale, D.A., Rapoport, A., 2000. Elicitation of strategy profiles in large group coordination games. *Exp. Econ.* 3, 153–179.
- Selten, R., 1978. The chain store paradox. *Theory Decis.* 9, 127–159.
- Selten, R., Stoecker, R., 1986. End behavior in sequences of finite prisoner's dilemma supergames: a learning theory approach. *J. Econ. Behav. Org.* 7, 47–70.
- Shah, A.K., Oppenheimer, D.M., 2008. Heuristics made easy: an effort-reduction framework. *Psychol. Bull.* 134 (2), 207–222.
- Smith, V.L., Suchanek, G., Williams, A., 1988. Bubbles, crashes and endogeneous expectations in experimental spot asset markets. *Econometrica* 56, 1119–1151.
- Stahl, D.O., 1996. Boundedly rational rule learning in a guessing game. *Games Econ. Behav.* 16, 303–330.
- Stahl, D.O., 1999. Sophisticated learning and learning sophistication. University of Texas at Austin Working Paper.
- Stahl, D.O., 2000. Local rule learning in symmetric normal-form games: theory and evidence. *Games Econ. Behav.* 32, 105–138.
- Stahl, D.O., Wilson, P., 1995. On players models of other players: theory and experimental evidence. *Games Econ. Behav.* 10, 213–254.
- Sundali, J.A., Rapoport, A., Seale, D.A., 1995. Coordination in market entry games with symmetric players. *Org. Behav. Human Decis. Processes* 64, 203–218.
- Sutton, R.S., Barto, A.G., 1998. *Introduction to Reinforcement Learning*. MIT Press, Cambridge.
- Tonin, M., Vlassopoulos, M., 2013. Experimental evidence of self-image concerns as motivation for giving. *J. Econ. Behav. Org.* 90 (C), 19–27.
- Van Damme, E., 1999. Game theory: the next stage. In: Gerard-Varet, L.A., Kirman, A.P., Ruggiero, M. (Eds.), *Economics Beyond the Millennium*, Oxford University Press, 184–214.

- Van Huyck, J., Battalio, R., Beil, R., 1990. Tacit cooperation games, strategic uncertainty, and coordination failure. *Am. Econ. Rev.* 80, 234–248.
- Van Huyck, J., Cook, J., Battalio, R., 1997. Adaptive behavior and coordination failure. *J. Econ. Behav. Org.* 32, 483–503.
- Von Neumann, J., Morgenstern, O., 1944. *Theory of games and economic behaviour*. Bull. Amer. Math. Soc. Princeton University Press, 18–625.
- Walker, M., Wooders, J., 2001. Minimax play at Wimbledon. *Am. Econ. Rev.* 91 (5), 1521–1538.
- Watson, J., 1993. A ‘reputation’ refinement without equilibrium. *Econometrica* 61, 199–205.
- Watson, J., Battigali, P., 1997. On ‘reputation’ refinements with heterogeneous beliefs. *Econometrica* 65, 363–374.
- Wooders, J., 2010. Does experience teach? Professionals and minimax play in the lab. *Econometrica* 78 (3), 1143–1154.
- Wright, J., Leyton-Brown, K., 2013. Evaluating, understanding, and improving behavioral game theory models for predicting human behavior in unrepeated normal-form games. Working Paper.
- Wright, N.D., Hodgson, K., Fleming, S.M., Symmonds, M., Guitart-Masip, M., Dolan, R.J., 2012. Human responses to unfairness with primary rewards and their biological limits. *Scientific Reports*, Wellcome Trust.
- Yamagishi, T., 1986. The provision of a sanctioning system as a public good. *J. Pers. Social Psychol.* 51 (1), 110–116.
- Zizzo, D.J., 2003. Empirical evidence on interdependent preferences: nature or nurture?, *Cambridge J. Econ.*, 27 (6), 867–880.