bates. In diplomatic decisions, such as whether to impose an embargo or call a bluff, knowing how another country's leader will behave is crucial. In designing incentive contracts, it is important to guess how workers will respond to incentives. In a review of empirical work, Prendergast (1999) noted that workers usually respond rationally (e.g., by working harder when a piece rate payment for output is imposed), but firms do not usually use contracts that are predicted by theory to work best. This broad pattern is consistent with companies maximizing profits but being unwilling to bet heavily that workers will optimize. In political science it is often assumed that people vote "strategically"—i.e., they will vote against their most preferred candidate if doing so can increase the utility they get from the likely outcome. Evidence suggests however that people often vote surprisingly "sincerely" (Alvarez and Nagler, 2002). In the 2000 U.S. presidential election, for example, 4% of people voted for Ralph Nader, who had no real hope of winning. Most of these voters strongly preferred Al Gore over George Bush. Gore would have won if the Nader fans had voted for him; because they didn't, Bush won, giving the Nader voters their worst outcome.

Knowing how many steps of iterated reasoning players use is also essential for giving good advice. In the beauty contest game, for example, the unique Nash equilibrium is zero. But actually choosing zero in an experiment is a mistake, because the goal is to be one step in reasoning ahead of others, *but no further*. Good advice depends on having a good idea of what others will do.

As we shall see below, the number of steps of iterated thinking people use is limited for several reasons. The assumption that *others* obey dominance has a very different cognitive status than strict dominance: It is a guess about the rationality of other players (and about what the other players' payoffs are). Although it seems quite reasonable for players to obey dominance, it is less obvious that you should always expect others to obey dominance. Iterating further, to think that others will think *you* obey dominance is a guess about what somebody else thinks about you. As players proceed up this hierarchy of iterated reasoning, the scaffolding they climb gets more and more wobbly. The psycholinguist Herbert Clark jokes that the grasp of three or more levels of iterated reasoning "can be obliterated by one glass of decent sherry" (Clark and Marshall, 1981).

The experiments establish empirical bounds on the number of steps of reasoning players use. This approach has been rapidly successful. The first papers along these lines are less than ten years old, but in that time-span regular behavior has been observed in a very wide range of games. What have we learned? It is important to note one definition and a qualification. I define obeying dominance to be *one* step of iterated deletion of dominated strategies. Thus, when I say "players seem to use one level of iterated dominance," that means they obey dominance but do not believe that others will obey dominance. The qualification is that game payoffs are defined as utilities, and therefore measuring whether people violated dominance requires measuring utilities. But we can control money payoffs in experiments and make guesses about how utilities depend on payoffs. Therefore, for the purposes of measuring dominance violations, I will assume that players' utilities depend only on their own payoffs. Otherwise, it is difficult to tell whether a player is violating dominance, per se, or exhibiting a utility for something other than her own payoffs (see Chapter 2).

Nearly all people use one step of iterated dominance (i.e., they obey dominance).[3] However, at least 10% of players seem to use each of two to four levels of iterated dominance, and the median number of steps of iterated dominance is two. A conclusion this sharp is of course a simplification; readers who want a more precise answer should keep reading.

The chapter is divided into five sections. Section 5.1 is a warmup which reports results from the simplest games, in which iterations of dominance eliminate one or two strategies, and a "patent race" game in which iteration eliminates some strategies but not others. Section 5.2 discusses the "*p*-beauty contest" games mentioned above, in which players choose numbers and the player whose number is closest to p times the average of the numbers wins a fixed prize. Section 5.3 discusses games economists have studied for various purposes, in which iterated applications of dominance *reduce* the collective payoffs to players: Centipede, I and S games (akin to centipede and prisoners' dilemma), price competition, travelers' dilemma, and email games. Section 5.4 is about the "dirty faces" game used in logic, which has the opposite property of the games in the third section—namely, an increase in the number of steps of iterated dominance *increases* payoffs. Section 5.5 is a betting game. Section 5.6 describes two ambitious studies which posit a set of "types" (who use different amounts of iterated reasoning, or various decision rules) and estimate the frequency of types statistically. Section 5.7 draws conclusions and suggests ways that game theory might be modified to account for limited iterated reasoning.

Here is some tour-guide advice: The basic regularities that appear again and again throughout the chapter—one to three steps of iterated reasoning—can be grasped from the simpler games in Sections 5.1 and 5.2. Understanding the games in Sections 5.3 and 5.4, and the econometrically intensive studies in Section 5.6, requires more knowledge of game theory

[3] There is an important exception to the rare-violation rule: Dominance violations are common when doing so is an expression of social preference. For example, cooperating in the prisoners' dilemma manifests cooperativeness; contributing in dictator games manifests generosity; paying back money in a trust game manifests trustworthiness; and rejecting ultimatum offers manifests vengeance, even though all these actions violate dominance (assuming own-payoff maximization).

and more patience. Pick and choose games in Section 5.3 to read about based on your tastes: The centipede game is easy to grasp and closely related to trust games, discussed in Chapter 2. The I and S games of Van Huyck et al. illustrate subtleties of trigger strategy logic in repeated games. The price competition game and traveler's dilemma are relatively simple. The email game is tricky but illustrates how learning can create a surprising degree of equilibration. Sefton and Yavaş's fine mechanism experiment illustrates a rare showdown between "high theory" and experimental observation.

5.1 Simple Dominance-Solvable Games

Several experiments measure behavior in games that can be solved with only two or three levels of iterated dominance.

5.1.1 Games Solvable by Two Steps of Iterated Dominance

The simplest study, and one of the earliest, was done by Beard and Beil (1994). Their study was motivated by an example of Rosenthal (1981) designed to question the descriptive accuracy of backward induction (or, in a normal form game, iterated dominance). Payoffs in the game (in dollars) are shown in Table 5.1.

Player 1 moves first and can earn \$9.75 for herself by choosing L (giving \$3 to player 2). Or she can choose R, putting player 2 on the move. If player 2 behaves self-interestedly she responds with r, giving the two players \$10 and \$5, respectively. If player 2 violates dominance by choosing l, they earn \$3 and \$4.75. Subgame perfection selects the solution (R,r), if player 1 is self-interested and thinks player 2 is self-interested also.

This game tests whether player 1 is willing to bet heavily that others will obey dominance. By varying the game payoffs, Beard and Beil tested for various influences on frequency of subgame perfection and on player 1's

Table 5.1. *Beard and Beil's iterated dominance game*

	Player 2 move	
Player 1 move	l	r
L	9.75, 3	
R	3,4.75	10,5

Source: Beard and Beil (1994).

Table 5.2. Payoff treatments and results in Beard and Beil

Treatment	Payoffs from (L, l)	(R, l)	(R, r)	Frequency of L	r|R	Number of pairs	Threshold $p(r\|R)$
1 (baseline)	(9.75,3)	(3,4.75)	(10,5)	0.66	0.83	35	.97
2 (less risk)	(9,·)	(·,·)	(·,·)	0.65	1.00	31	.85
3 (even less risk)	(7,·)	(·,·)	(·,·)	0.20	1.00	25	.57
4 (more assurance)	(·,·)	(·,3)	(·,·)	0.47	1.00	32	.97
5 (more resentment)	(·,6)	(·,·)	(·,·)	0.86	1.00	21	.97
6 (less risk, more reciprocity)	(·,5)	(5,9.75)	(·,10)	0.31	1.00	26	.95
7 (1/6 payoff)	(58.5,18)	(18,28.5)	(60,30)	0.67	1.00	30	.97

Source: Beard and Beil (1994).
Note: (·,·) indicates the payoffs are the same as those in the baseline case.

beliefs about 2's likelihood of violating dominance. Table 5.2 summarizes the various payoffs and results (dots indicate the same payoffs as in the baseline treatment 1).

In the baseline treatment, 66% of player 1s choose L, while 83% of the R choices are met with the self-interested response r. The faith in 2's rationality required to justify choosing R is shown by the threshold probability $p(r|R)$; this is the belief in an r choice following R which makes R just preferable to L (if she is risk neutral). The threshold is 0.97 in the baseline treatment. Since 83% of player 2s choose r, the threshold is not quite met by actual behavior.

The "less risk" treatments 2–3 lower the risk of choosing R by lowering the L payoff to player 1. (The thresholds $p(r|R)$ fall to 0.85 and 0.57.) Player 1s then choose L less often—65% and 20%. In "more assurance" treatment 4, the gap in player 2's R payoffs is raised, providing more incentive for 2 to choose r instead of l. Player 1s respond by choosing L less often. In "more resentment" treatment 5, the payoff to player 2 from L is raised from \$3 to \$6. This is designed to create resentment in player 2 if 1 chooses R, which forces 2 to accept less than she would have gotten if L had been chosen. This treatment raises the fraction of nonsubgame perfect play of L to its highest level, 86%. The "more reciprocity" treatment again lowers risk for player 1, by raising player 2's R payoffs; then player 2 might feel inclined to reciprocate a "generous" choice of R by choosing r. This treatment lowers L choice to 31%. Finally, paying large stakes probabilistically, in treatment 7, does not affect the results at all relative to the baseline treatment 1 with equivalent expected payoffs.

Table 5.3. *Goeree and Holt's credible threat games*

Condition	Number of pairs	Threshold $p(r\|R)$	Payoffs (L)	(R,l)	(R,r)	Frequency of L	r|R
Baseline 1	25	.33	(70,60)	(60,10)	(90,50)	0.12	1.00
Lower assurance	25	.33	(70,60)	(60,48)	(90,50)	0.32	0.53
Baseline 2	15	.85	(80,50)	(20,10)	(90,70)	0.13	1.00
Lower assurance	25	.85	(80,50)	(20,68)	(90,70)	0.52	0.75
Very low assurance	25	.85	(400,250)	(100,348)	(450,350)	0.80	0.80

Source: Goeree and Holt (1999).

As Rosenthal conjectured, 1s do not usually have enough faith in 2's rationality and self-interest to choose the subgame perfect choice R, except when the assurance threshold is around a half (treatment 3) or player 2 seems likely to reciprocate a generous choice (treatment 6). At the same time, 2 *does* choose the self-interested response r in almost every case. Player 2s overwhelmingly obey dominance but player 1s are not willing to bet that player 2s will obey dominance.

The basic finding of Beard and Beil (1994) was replicated by Goeree and Holt (1999), in their insightful paper on "treasures and contradictions" in game theory.[4] Table 5.3 shows their results. The fraction of L moves varies from 12% to 80%. As in Beard and Beil's study, the tendency to play L responds predictably to differences in player 1's risk, and to the incentive player 2 has to play r rather than l. Starting from baseline condition 1, lowering the assurance that player 2 will move r (by making the differences between 2's payoffs in (R,r) and (R,l) closer) lowers the fraction of player 2s who *actually* move r, and raises the propensity of player 1s to move L. The same effect occurs starting from baseline 2. The *cost* of player 2's deliberate punishment or mistake affects what player 2 does, and player 1s seem to anticipate this in deciding whether to take the safe action L.

5.1.2 Iterated Dominance and Tree-Matrix Differences

Schotter, Weigelt and Wilson (1994) did a more extensive comparison of iterated dominance and subgame perfection. Their first games, 1M and 1S, are shown in Table 5.4. Games 1M and 1S are like the Beard and Beil games,

[4] The theme of their paper is that pairs of games which have the same kinds of equilibria can be constructed so that players choose the equilibrium in one game (treasure) but do not choose it in the equivalent paired game (contradiction). The major difference is that Goeree and Holt use the strategy method: They asked player 2s what they would do *if* player 1 chose R.

Table 5.4. *Games 1M and 1S of Schotter et al.*

Player 1	Player 2 l	r	Actual frequency
Normal form (1M)			
L	4,4	4,4	(0.57)
R	0,1	6,3	(0.43)
Frequency	(0.20)	(0.80)	
Sequential form (1S)			
L	4,4		(0.08)
	l	r	
R	0,1	6,3	(0.92)
Frequency	(0.02)	(0.98)	

Source: Schotter, Weigelt, and Wilson (1994).

a test of player 1's willingness to expect 2 to obey weak dominance (game 1M) or play self-interestedly in the subgame (game 1S).

A large majority of player 2s *do* obey dominance—80% in 1M and 98% in 1S—but player 1s are willing to bet strongly on this (choosing R) only in the sequential game 1S. The difference in 1M and 1S behavior seems to be due to a matrix-tree "presentation effect."[5]

Schotter et al. also studied a hybrid game (1H) in which the game was described sequentially—players were actually shown a tree—but *played* simultaneously. In this hybrid version the fractions of R and r play were 86% and 88%, similar to results in the sequential version 1S. It appears that the physical description of the game is what matters. Perhaps the visual isolation of player 2's move in the tree makes dominance of r over l more transparent to player 1.

Their games 3M and 3S, in Table 5.5, allow investigation of three levels of iterated dominance (game 3M) and forward induction (game 3S). In 3M, B is strictly dominated for player 1. Eliminating it makes M weakly dominant for player 2. Eliminating T and B for player 2 then selects M for player 1, so (M,M) is selected by three steps of iterated dominance.

In an equivalent sequential game, 3S, player 1 can move T and end the game with payoffs (4,4), or put player 2 on the move. Player 2 can then

[5] Another example is the common violation of weak dominance manifested in sealed-bid second-price auctions, compared with strategically equivalent English ascending-price auctions (see Kagel, 1995; Camerer, 1998).

Table 5.5. Games 3M and 3S of Schotter et al.

	Player 2 move					
Player 1 move	T	M	B	Frequency		
Normal form 3M						
T	4,4	4, 4	4,4	(0.82)		
M	0,1	6,3	0,0	(0.16)		
B	0,1	0,0	3,6	(0.02)		
Frequency	(0.70)	(0.26)	(0.04)			
Sequential form 3S						
						Conditional frequency
T	4,4					(0.70)
		T				
		0,1				
				M	B	
			M	6,3	0,0	(1.00)
			B	0,0	3,6	(0.00)
Frequency		(0.13)		(0.31)	(0.69)	

Source: Schotter, Weigelt, and Wilson (1994).

end the game with T, yielding (0,1), or else they play a simultaneous-move battle-of-the sexes (BOS) game with pure-strategy equilibria (M,M) and (B,B). In 3S, iterated dominance eliminates strategies in conjunction with an assumption of dynamic consistency, leading to (M, M). T is dominated for player 2 since she can guarantee more than 1 by mixing in the BOS subgame (if she is not too risk averse). And forward induction applies: If player 2 realizes that player 1 rejecting the (4,4) payoff signals 1's intention to play *M* in the subgame and get 6, then player 2 will also play M. This argument picks out the (M,M) equilibrium. The crucial difference in the game forms is that player 2 in 3S can move after *observing* what player 1 has done, rather than merely hypothesizing what 1 would do in game 3M (and 2 knowing in 3M that 1 realizes 2 is merely hypothesizing rather than "knowing").

Table 5.5 shows that in game 3M hardly any player 1s violate strict dominance by choosing B, but only a small minority of player 2s (26%) seem to anticipate that and deduce that they should play the weakly dominant

M. Thus, there is little evidence for more than one step of deletion of dominated strategies.[6]

The sequential 3S results are roughly similar. In the subgame player 1s always choose M. However, player 2s do not seem to figure this out and mistakenly choose B 69% of the time. Apparently anticipating this, player 1s mostly choose T, so the forward induction equilibrium (M,M) is rarely reached.

Overall, Schotter et al. saw very limited evidence of either much iteration of dominance (beyond one step), or subgame perfection and forward induction, except in the simplest case 1S. They speculate that more experience is needed to lead to dominance-solvable outcomes. Brandts and Holt (1995) did experiments with eight periods, however, and still observed only limited evidence of forward induction.

5.1.3 A Partially Dominance-Solvable Patent Race Game

Rapoport and Amaldoss (1997) ran a "patent race" investment game in which iterated deletion has an interesting alternating structure (see Zizzo, 2002). In their game the players are "strong" and "weak" (or deep- and shallow-pocketed firms). The strong player has an endowment of 5 and the weak player has an endowment of 4. Whichever player spends the most earns a reward of 10 (independent of what they spent), and keeps her endowment minus what she spent. If both players spend the same amount, neither one gets the reward. The payoffs are shown in Table 5.6.

Denote strong player investments of i by s_i and weak player investments by w_i. The stronger player can guarantee earning the reward by investing all 5, choosing s_5, simply outspending the weak player. Since this strategy strictly dominates investing 0 (and keeping the endowment of 5), strategy s_0 can be eliminated for the strong player. Eliminating s_0 then makes investing 1 dominated for the weak player (since investing 1 is only worthwhile if she has a chance of outspending the strong player, and she can't if the strong player never invests 0). Eliminating w_1 for the weak player then makes s_2 dominated by s_1 for the strong player, and so forth. Applying dominance iteratively leads to deletion of strategies in the following order: s_0, w_1, s_2, w_3, s_4. Further analysis derives predicted mixed-strategy probabilities for the undeleted strategies: Strong players should choose the highest investment, s_5, 60% of the time, and the weak players should throw in the towel by choosing the

[6] Schotter et al. also ran games 3Mb and 3Sb in which they changed the (0,0) payoffs to (2,2). This does not change the game-theoretic prediction at all but has one effect on the results. Player 2s in 3Mb chose T, M, and B 11%, 32%, and 57% of the time, respectively, rather than 70%, 27%, and 4%. Changing the 0 payoff from player 2 strategies M and B to a payoff of 2 makes the fact that T is dominated more obvious and reduces its choice frequency from 70% to 11%.

Table 5.6. Payoffs in patent race game

Weak player investment, w_i	Strong player investment, s_i 0	1	2	3	4	5	Prediction	Actual frequency
0	4,5	4,14	4,13	4,12	4,11	4,10	0.60	(0.55)
1	13,5	3,4	3,13	3, 12	3,11	3,10	0.00	(0.03)
2	12,5	12,4	2,3	2,12	2,11	2,10	0.20	(0.07)
3	11,5	11,4	11,3	1,2	1, 11	1,10	0.00	(0.14)
4	10,5	10, 4	10, 3	10,2	0,1	0,10	0.20	(0.22)
Prediction	0.00	0.20	0.00	0.20	0.00	0.60		
Frequency	(0.01)	(0.17)	(0.05)	(0.09)	(0.13)	(0.55)		

Source: Rapoport and Amaldoss (1997).

lowest investment, w_0, 60% of the time. Both should randomize equally over their other two undeleted strategies.

The predicted probabilities and actual frequencies across 160 periods are shown in Table 5.6. Conformity to this unintuitive equilibrium is quite good. Players were predicted to choose extreme investment levels 60% of the time and actually chose them 55% of the time. The fractions of play of the iteratively dominated strategies, in the order in which they are eliminated, were 0.01, 0.03, 0.05, 0.14, and 0.13. At the individual level, the fractions of players who violated one or more levels of iterated dominance at least once in 80 trials were 0.11, 0.32, 0.83, 0.70, and 0.92 (in order of increasing number of steps of iteration).[7]

Summary: In the simplest games in which iterated application of dominance deletes some strategies, few subjects violate dominance, but most subjects are also not willing to bet heavily that others will obey dominance (i.e., there is one step of iterated dominance). In the patent race games, most subjects exhibit three levels of iterated dominance, but some of their sophisticated behavior might be due to learning over 160 trials.

5.2 Beauty Contest Games

The "*p*-beauty contest" game first presented in Moulin (1986), and discussed earlier, is an ideal tool for measuring the number of steps of iterated deletion of dominated strategies. Each of N players i choose a number x_i in the

[7] That is, 11% of the strong players played s_0 at least once (89% never did), 32% of the weak players played w_1 at least once, and so on.

interval [0,100] simultaneously. A multiple p of the average of their numbers, $p \cdot \sum_{i=1}^{N} x_i/N$, defines a target number. The player whose number is closest to the target number wins a fixed prize. Before proceeding, readers should think of which number they would pick if they were playing against a group of students.

The game is called a "beauty contest" after the famous passage in Keynes's *General Theory of Employment, Interest, and Money* about a newspaper contest in which people guess which faces others will guess are most beautiful (see Chapter 1). Like people choosing the prettiest picture in Keynes's passage, players in the beauty contest game must guess what average number others will pick, then pick 2/3 of that average (knowing everyone is doing the same). The beauty contest game can distinguish whether people "practise the fourth, fifth, and higher degrees" of reasoning as Keynes wondered. Choosing a number larger than 67 violates stochastic dominance, because any such choice has less chance of winning than a choice of exactly 67 does. So numbers in the range (67,100] violate first-order iterated dominance. A player who thinks others obey dominance can infer that the target will be below (2/3)67, or 44, so an optimal choice is in the range [0,44]. Hence, a choice between (45,67] is consistent with a player obeying one step of dominance, but not two. Stepping further, choices in the range (29,44] are consistent with two steps of iterated dominance but not three. Thus, number choices in the beauty contest game place bounds on the frequency of violations of increasing degrees of iterated rationality. Infinitely many steps of iterated dominance lead to the unique Nash equilibrium of 0.

This game was first studied experimentally by Nagel (1995). She used groups of fourteen–sixteen German students as subjects. Her results from games with $p = 2/3$ are shown in Figure 5.1b. The average number is around 35, and many subjects chose either 33 (one step of reasoning from the midpoint of 50) or 22 (two steps). Very few subjects picked 0.

The first replication of Nagel's results was reported by Ho, Camerer, and Weigelt (1998). They used values of p of 0.7, 0.9, 1.1, and 1.3, to compare behavior when the Nash equilibrium is 0 (when $p < 1$) and the Nash equilibrium is the largest number (which results when $p > 1$). Subjects played one game with a particular value of p 10 times, then played another game in which p was on the opposite side of 1, so the equilibria in the two games were on opposite ends of the number interval.[8] There was an interesting effect of transfer of learning across the two games: Subjects who

[8] Subjects were students in Singapore in groups of three or seven playing for $3.50. Curiously, subjects in the larger groups tend to choose numbers closer to equilibrium and converge faster. We suspect that in the three-person groups, subjects think about each of the other two subjects separately, and generally want to be in the middle. In the seven-person groups, they think about the composite of the six other subjects, and want to be below the composite, which drives them closer to zero. Quantal-response equilibrium also predicts lower choices when groups are larger.

Table 5.7. *Estimated fractions ω_k of level-k types in beauty contest games*

	Ho, Camerer, and Weigelt (1998)		Nagel (1995)	
Estimate	$p > 1$ games	$p < 1$ games	$p = 1/2$	$p = 2/3$
ω_0	0.22	0.16	0.16 (0.24)	0.28 (0.13)
ω_1	0.31	0.21	0.38 (0.30)	0.34 (0.44)
ω_2	0.13	0.13	0.47 (0.41)	0.37 (0.39)
ω_3	0.34	0.50	0.00 (0.06)	0.00 (0.03)

Note: Numbers in parentheses indicate Nagel's original estimates.

had converged toward a high equilibrium in one game (with $p > 1$) tended to start higher in the second game with $p < 1$, but also converged more rapidly, as if they "learned to learn" from one game to another (see Camerer, Ho, and Chong, 2002a,b, and Chapter 6).

In further experiments, Ho, Weigelt, and I became interested in the influence of the size of the stakes on behavior. Higher stakes might induce subjects to think harder or might lead subjects to think *others* would think harder, which would lead them to choose lower numbers. We compared low (\$7) and high (\$28) stakes conditions across ten periods; the results were shown in Figures 1.3a and 1.3b. There was a small effect of stakes lowering number choices (i.e., increasing the number of steps of thinking), especially in later periods when virtually all the high-stakes subjects chose numbers less than 1.

In our paper we also improved on the casual way in which Nagel measured steps of iterated reasoning. Following Stahl and Wilson (1995) (see Section 5.5 below), we assumed that a fraction ω_0 of the subjects just choose a number randomly from a normal distribution with mean μ and standard deviation σ. Call these subjects "level-0" players. A fraction ω_1 are level-1 players, who think all others are level-0 players and choose best responses with some noise. There are also assumed to be fractions ω_2 and ω_3 of level-2 and level-3 players, each of whom thinks all others are one level below them and best-respond.

Table 5.7 shows the estimated fractions of the various levels of players, using only first-round data. The estimated fractions of levels of reasoning show that players are typically using one to three steps of iterated reasoning. Nagel's informal estimates (shown in parentheses) are close to the estimates derived by our earlier procedure.

Several variants of the beauty contest game have been done. Figure 5.1 shows histograms taken from Nagel (1999). Using the median instead of the average to compute the target does not change results (Figure 5.1c).

Figures 5.1g–h show experiments by Ho, Weigelt, and myself (unpublished) in which the equilibrium is located in the *interior* of the number

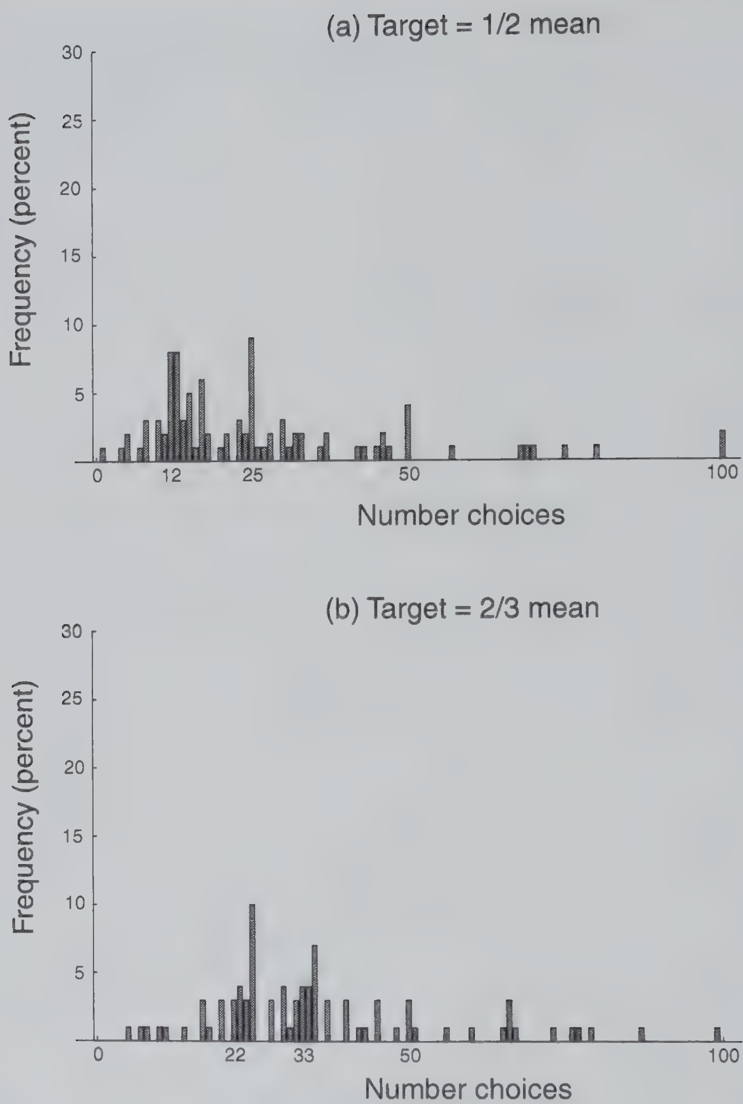
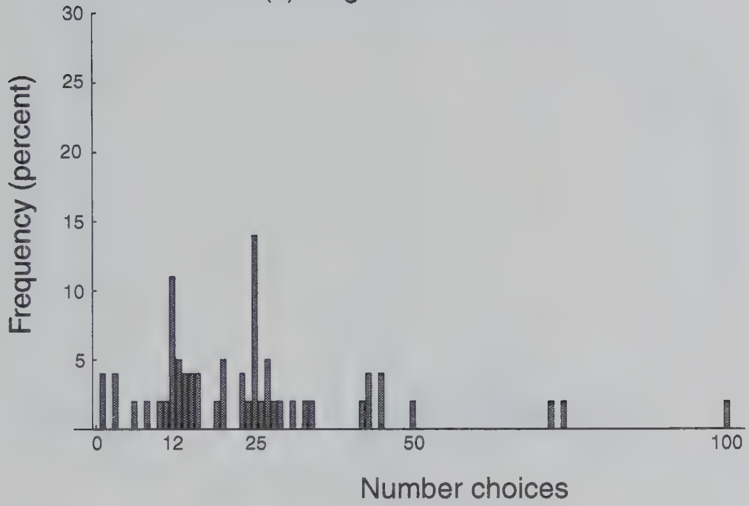


Figure 5.1. Beauty contest choices. Sources: Based on data from Nagel (1999) and Camerer, Ho, and Weigelt (unpublished).

(c) Target = 1/2 median



(d) Target = 2/3 mean, game theorists and experimenters

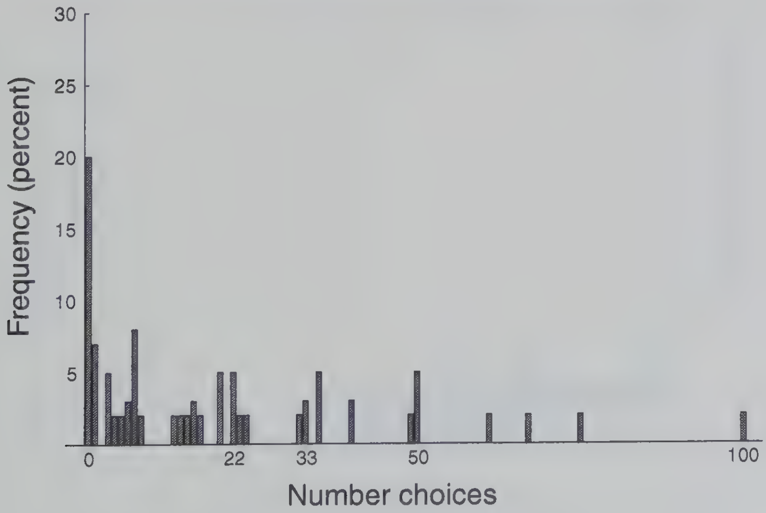
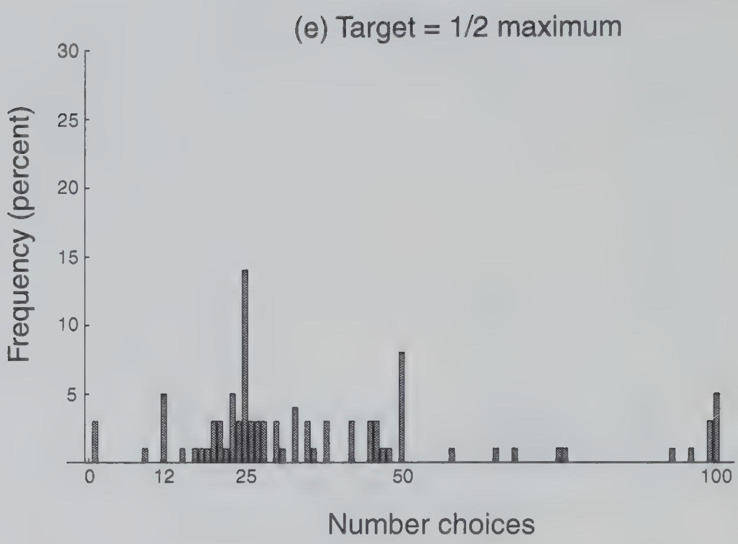


Figure 5.1 (continued)



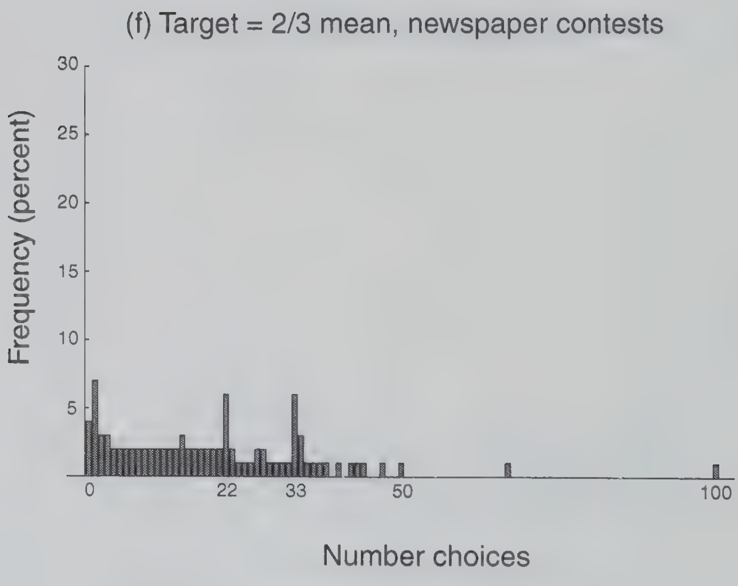
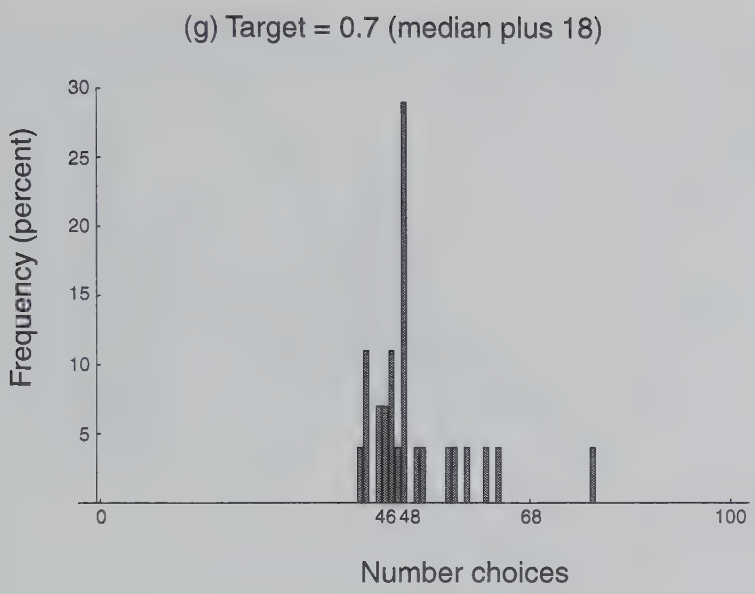


Figure 5.1 (continued)



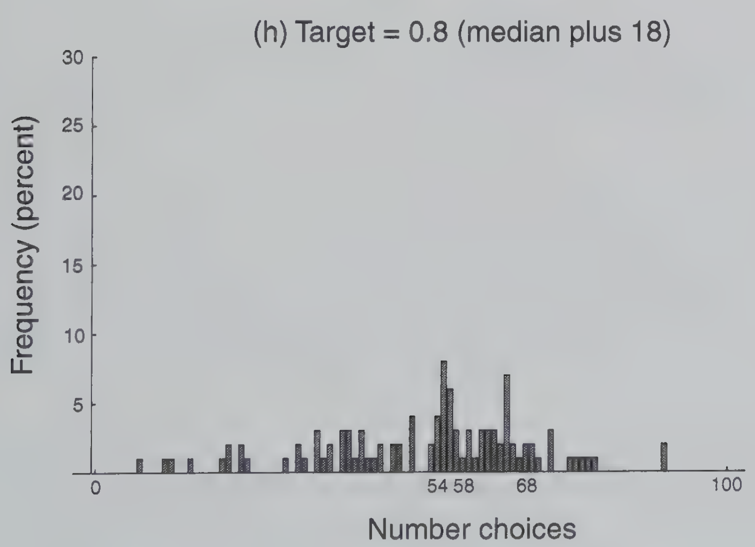


Figure 5.1 (continued)

interval. When the target is $p(M+18)$ (where M is the median number), then the equilibrium is 42 if $p=0.7$ and 72 if $p=0.8$. Locating the equilibrium in the interior is useful because having an equilibrium on the boundary makes it impossible, strictly speaking, to distinguish equilibrium play plus random noise from limited iterated reasoning.

Nagel's graphs helpfully mark the x-axis with numbers that correspond to various steps of reasoning (starting from 50). For example, in the $M+18$ game with $p=0.8$, subjects who think others will choose an average of 50 and best-respond will choose 54; best-responding to 54 yields 58, and so forth. The distributions show spikes at numbers that correspond to one or two steps of reasoning.

The beauty contest game has been used in demonstration experiments from many exotic subject pools in public lectures. These data have less experimental control than in the laboratory, but they are informative about whether unusual groups who we cannot usually get into the lab behave like college students.

Figure 5.2 shows data from several unusual subject pools (see Camerer, 1997). Caltech students are in the front of the figure, followed by economics Ph.D.s, and so on, with portfolio managers last. One group is portfolio managers who make investment decisions on behalf of large groups of people (and who are savvy about how stock markets work). Another group is students in a graduate (Ph.D.) course in economics, who have all learned about simple game theory concepts. Still another group is Caltech undergraduates in a psychology course, using $p=2/3$. Their average was around 24, about a half-step of reasoning lower than in other student populations. These data are useful because it is often asserted that subjects who are "smart enough" will make choices much closer to game-theoretic predictions than average college subjects will. The median SAT math score at Caltech is often 800, the maximum. The fact that the Caltech students do not choose numbers that much closer to the Nash equilibrium than average folks refutes the hypothesis that simply being good at math will automatically lead players to a Nash equilibrium.

Giving a lecture to the Caltech Board of Trustees, I had a chance to see how older adults with an amazingly wide range of personal achievements would play this game in a group of their peers. The group had a subsample of twenty CEOs, corporate presidents, and board chairmen. Members of this subsample are indeed among the titans of industry who are often thought of as the "rational decision makers" whose ideas and behavior influence the entire economy. Their results are shown in the CEO histogram in Figure 5.2.

Another interesting subject pool is readers of financial magazines. Two experiments were conducted, one by Rosemarie Nagel and Antonio Bosch in the Spanish business magazine *Expansion*, another by Richard Thaler in the British newspaper *The Financial Times*. Readers were offered a large prize

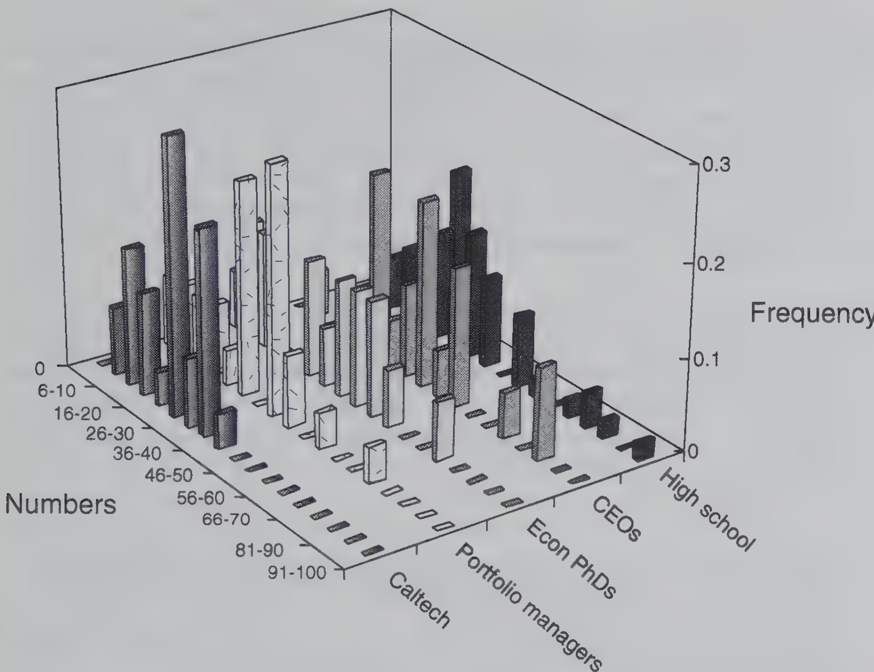


Figure 5.2. *Beauty contest choices from various exotic subject pools. Source: Camerer (unpublished).*

if they won a $p = 2/3$ contest. The pooled results from these two newspaper contests are shown in Figure 1f. There is a sharp spike at 33 and another spike at 22. About 8% of subjects chose the equilibrium of 0. Interestingly, there is no prominent spike between 22 (two steps of reasoning) and 0. Players who did more than two steps of reasoning seemed to pick up momentum rolling down a slippery inductive slope, and continued their logical induction all the way to 0.[9]

Slonim (2001) had subjects play three beauty contests for three periods each, and mixed together players with different levels of experience. When an experienced player is suddenly mixed in with two new subjects who have not played before, what happens? One possibility is that the experienced

[9] It is also notable that the *Financial Times* and *Expansion* readers had a lower average number, and many fewer outlying responses above 50, than subjects in other experiments. This fact is, frankly, a little disturbing for experimenters because it points out the way in which a highly educated group of subjects, volunteering to participate in a mass public experiment, might reason more thoughtfully than students corralled together to make a choice in the usual experimental style.

player recalls what *she* did when she first played the game, and is able to use that information to outguess the newbies. Or it might be that choosing low numbers has become so natural or obvious to the experienced player, that she chooses *too low* a number (adjusting too little for the fact that new players will pick high numbers, a "curse of knowledge") and does *worse* than the new folks. It turns out that the experienced players use their experience wisely—they choose higher numbers when the new players arrive, and win almost all the time. But their edge disappears after one period.

Summary: In beauty contest games, the typical subject uses one or two steps of reasoning (starting from 50). This basic result has been obtained with the widest variety of subject pools ever sampled for a particular game. Unusual analytical skill and training in game theory move choices about one iteration closer to the equilibrium. Samples of self-selected newspaper readers also make choices closer to equilibrium.

5.3 Games in Which Iterated Reasoning Decreases Payoffs

This section discusses dominance-solvable games in which higher levels of iterated reasoning reduce the collective payoffs to players.

5.3.1 Centipede Games

By repeatedly passing, the players can enlarge the pie dramatically, but as the pie grows the temptation to choose Take does too.

A four-move "centipede" game[10] studied by McKelvey and Palfrey (1992) is shown in Figure 5.3. At each node, a player can "take" (T) 80% of a growing pie and end the game, or "pass" (P) and let the other player move, doubling the pie each period. The collective gain to passing is huge. In the four-move game, players could share \$8 (\$6.40 + \$1.60) if they pass at every node, but if a player takes immediately they share just \$0.50.

Centipede games are multistage trust games (Chapter 2 discusses one-stage trust games). They model situations such as business or personal relationships in which there are gains from exchange which grow over time (perhaps owing to learning), and a constant temptation for each player to end the relationship to grab more. Backward induction (along with self-interest) predicts that players will expect taking at the last node (since passing at that point violates dominance), which leads to taking at the next-to-last node, and so forth to the start: Players will take at *every* node. The

[10] The game (introduced by Rosenthal, 1981) was dubbed a "centipede" game by Ken Binmore because it has one hundred vertical terminal nodes which, graphically, resemble a centipede insect.

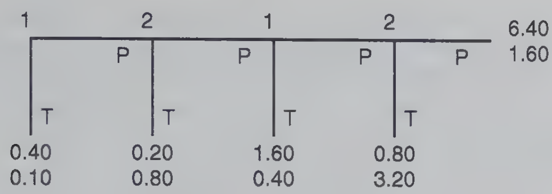


Figure 5.3. A centipede game with four moves. Source: Based on McKelvey and Palfrey (1992).

game measures steps of iterated reasoning because passing at node t violates $5 - t$ steps of iterated dominance.

The centipede game illustrates an important byproduct of experimentation: Choosing a design disciplines theorizing because it forces one to be crystal clear about the conditions under which theory is really expected to apply. In centipede games, the dollar payments to subjects vary dramatically across possible outcomes. In a six-move game, the Nash equilibrium predicts the experiment will cost $.50 per trial. But if subjects completely trust each other and pass to the end it will cost $32, sixty-four times as much! Choosing the scale of payoffs requires an experimenter literally to bet money—the experimental budget—on what will happen.

The baseline game is the four-move game shown in Figure 5.3. McKelvey and Palfrey used undergraduate subjects at Caltech (CIT) and Pasadena City College (PCC), which spans a range of analytical skill. They also ran a six-move game and a version of the four-move game with payoffs multiplied by four. Table 5.8 shows results aggregated across four PCC and two CIT sessions for sequences 1–5 and 6–10.

Table 5.8. Results in McKelvey and Palfrey's centipede games

		Conditional frequencies of "take" at nodes					
Treatment	Trials	1	2	3	4	5	6
Four moves	1–5	0.06	0.32	0.57	0.75		
	6–10	0.08	0.49	0.75	0.82		
Four moves,	1–5	0.08	0.46	0.60	0.80		
high stakes	6–10	0.22	0.41	0.74	0.50		
Six moves	1–5	0.00	0.06	0.18	0.43	0.75	0.81
	6–10	0.01	0.07	0.25	0.65	0.70	0.90

Source: McKelvey and Palfrey (1992).

Players rarely take in early nodes as Nash equilibrium (and four steps of iterated dominance) predict. However, conditional "take" probabilities *do* increase at each node and most subjects (around 80%) take at the last node. There is some evidence of "unraveling" toward the equilibrium, because take probabilities are higher at every node in trials 6–10 than in the earlier trials 1–5. There is little effect from quadrupling the stakes.

By design, the subgame consisting of the last four moves in the low-stakes condition (beginning with the third node) is *exactly* the same as the four-move game with high stakes. If players build up trust over time, then they should be more likely to pass at the third node of the longer six-move game than at the start of the four-move game. Actually, the opposite is true. Having had two periods of prior trust in the six-move game does not enhance tendency to pass.

Following the insight of the famous "gang of four" paper on the repeated prisoners' dilemma (Kreps et al., 1982), McKelvey and Palfrey explain the high rates of passing with an equilibrium model in which a small percentage (7%) of altruistic players truly prefer to pass, and a large percentage of normal players mimic the altruists by passing up to a point in order to earn more.[11] Gradual unraveling is explained by including an error rate that shrinks over time (see also Zauner, 1999).

Fey, McKelvey, and Palfrey (1996) control for the influence of social preferences using a constant-sum centipede game. The players divide $3.20. At the first take node, the sum is divided evenly; as players pass, the division of the $3.20 gets more and more lopsided. If players pass in the regular centipede game just to create more surplus to share (as if conspiring against the experimenter), they will take right away in the constant-sum version of the game. In fact, players do take much more often. In the last half of the experiment, the conditional rates of playing take are more than half at the first node and around 80% at the second node. A quantal response equilbrium (QRE) model with an increasing response sensitivity (a reduced-form model of learning) fits the data reasonably well.

Nagel and Tang (1998) ran normal-form centipede game experiments with extensive-form feedback: Players state the first node at which they would "take"; the player who took at the later node is told the node at which the other player took but not vice versa (i.e., if you took first you don't know when the other player would have taken). Players take about halfway through the game, and there is no evidence of learning to take earlier. In fact, there is a slight movement toward taking *later*, away from the Nash equilibrium (see Ho, Wang, and Camerer, 2002).

[11] This idea was first informally implemented by Camerer and Weigelt (1988; see Chapter 8) and Palfrey and Rosenthal (1988), but McKelvey and Palfrey deserve credit for the first high-tech implementation, which marked the start of more sophisticated "experimetrics."

Table 5.9. *Centipede-type game I of Van Huyck et al.*

Row strategies	Column strategies				
	a_1	a_2	a_3	a_4	a_5
a_1	7,7	0,11	0,0	0,0	0,0
a_2	11,0	6,6	0,10	0,0	0,0
a_3	0,0	10,0	5,5	0,9	0,0
a_4	0,0	0,0	9,0	4,4	0,8
a_5	0,0	0,0	0,0	8,0	3,3
First period of repeated game I(5/6,2)					
First sequence	0.20	0.20	0.40	0.18	0.03
Last sequence	0.30	0.23	0.23	0.23	0.03

Source: Van Huyck, Wildenthal, and Battalio (2001).

Rapoport et al. (in press) ran three-person centipede games with enormous stakes (subjects could have made thousands of dollars if they passed to the end . . . but they didn't). They found a very low rate of passing. In a low-stakes condition, subjects lend more often, so a sufficient condition for Nash behavior seems to be three players and high stakes. (The cognitive hierarchy model of Camerer, Ho, and Chong, 2001, can explain the difference between two- and three-player games but QRE cannot.)

5.3.2 Prisoners' Dilemma and Quasi-Centipede Games

Van Huyck, Wildenthal, and Battalio (2002) investigate two 5×5 dominance-solvable games, shown in Tables 5.9 and 5.10. One has a centipede structure (game I) and one is a multi-strategy prisoners' dilemma (game S). The games are denoted I and S because I is solved by *i*terated dominance and S is solved by *s*trict dominance.

In game I, strategy a_2 weakly dominates[12] a_1; eliminating a_1 makes a_2 dominated by a_3, and so forth, until only a_5 is left. However, the unique Nash equilibrium (a_5, a_5) is Pareto dominated by any other (non-equilibrium) outcome on the diagonal.

Game S is more straightforward: Strategy a_5 strictly dominates all other strategies, leading to a unique Nash equilibrium (a_5, a_5) which is Pareto dominated.

[12] Lower-numbered strategies can also be eliminated by strict dominance using the mixed strategy which puts 0.70 weight on a_2 and smears positive support on each of the higher-numbered strategies.

Table 5.10. Prisoners' dilemma–type game S of Huyck et al.

Row strategies	Column strategies				
	a_1	a_2	a_3	a_4	a_5
a_1	7,7	0,0	0,0	0,0	0,11
a_2	0,0	6,6	0,0	0,0	0,10
a_3	0,0	0,0	5,5	0,0	0,9
a_4	0,0	0,0	0,0	4,4	0,8
a_5	11,0	10,0	9,0	8,0	3,3
First period of repeated game S(5/6,2)					
First sequence	0.25	0.05	0.08	0.00	0.63
Last sequence	0.65	0.00	0.00	0.00	0.35

Source: Huyck, Wildenthal, and Battalio (2001).

Van Huyck et al. were interested in the emergence of repeated-game equilibria when the game is played with a fixed-partner protocol. They compare treatments $G(\delta, T)$, where δ is the probability of continuing the game after each period (with the same partner), and T is the length of the terminal "continuation game" after the probabilistically-repeated game ends. Specifically, they compare one-shot games I(0,1) and S(0,1) with I(5/6,2) and S(5/6,2). In the latter games, players are told that there is a 5/6 chance the game will continue after each trial. When that phase ends, they play exactly two more periods, hence the notation (5/6,2). In the repeated game, the 5/6 continuation probability is high enough that it is an equilibrium for both players to choose a_1 and to punish a defection from this equilibrium by choosing the next-best repeated-equilibrium payoff.[13]

There is a subtle strategic difference between repeated games I(5/6,2) and S(5/6,2). Axelrod (1985) pointed out that trigger strategies—such as punishing one's opponent for defecting in a repeated PD—are "clear" in the sense that there is little doubt the punishment is meant to exact revenge for an earlier defection. In game I(5/6,2), playing a_2 is itself a repeated-game equilibrium *and* is also the optimal way to defect, in the short run, from the a_1 equilibrium. To see this, suppose in game I that we are bumping along happily playing a_1, and earning 7 each time. Now one player switches to a_2. Are they "defecting," which merits punishment by reverting

[13] In the PD-like game S, the next-best payoff is 11 from reverting to the Nash equilibrium a_5. In the centipede-like game I, the next-best payoff is the 11 from a_2. These conditions give long-run payoffs to choosing a_1 of $7 + \delta/(1-\delta)(7)$ and a payoff to defecting of $11 + \delta/(1-\delta)(3)$. For $\delta = 5/6$, sticking to the a_1 equilibrium yields a higher payoff than defecting (42 rather than 26).

to the Nash equilibrium a_5, or are they switching to the a_2 equilibrium? You simply can't tell. The PD game S(5/6,2) is different. In that game, if players are choosing a_1 reliably, and one player chooses a_2, that choice is not a profitable defection (since it yields a payoff of 0). Instead, a choice of a_2 is "clearly" a switch to the slightly less efficient repeated-game path of a_2. Van Huyck et al. argue that confusion about whether a defection in I(5/6,2) to a_2 is a punishment, or a reversion to a new equilibrium, may undermine the effectiveness of such a punishment as a way of supporting the Pareto-efficient repeated-game equilibrium at a_1. If they are right, there will be more choices of a_1 in S(5/6,2) than in I(5/6,2).

In the centipede-like game I(0,1), subjects paired randomly in one-shot games started out choosing a mixture of distributions with a median of a_3, and converged in about twenty rounds to a modal choice of a_5. Since the median a_3 corresponds to three steps of iterated dominance, this initial play of a_3 is roughly consistent with the other games in this chapter. In the one-shot PD-like game S(0,1), subjects overwhelmingly chose the dominant strategy a_5 from the beginning, although a few persistently chose the cooperative outcome a_1.

Behavior in the repeated games I(5/6,2) and S(5/6,2) is interesting. Some summary statistics are reported in the bottom panel of Tables 5.9 and 5.10. The statistics show the frequency distributions across actions for the *first* period in each sequence, averaged over the first and last sequences in each of the four sessions. These statistics show how subjects start off each time they play a separate sequence, and how this initial behavior varies across sequences.

Game I(5/6,2) results look like those in the one-shot game I(0,1): Subjects play a_1 through a_4 with almost equal probability. These frequencies do not change much from the first of the eight sequences in each session to the last sequence. Subjects are not learning, across sequences, to coordinate on the more efficient repeated-game equilibrium a_1.

Behavior is quite different in the repeated game S(5/6,2). In the first sequence more than half of subjects play the dominated strategy a_5. By the last (eighth) sequence, more than half are choosing a_1 in the first period. In the last five of the eight sequences, the median subject plays a_1 until the continuation phase is over, then reverts to a_5 in both periods of the terminal phase. Just as Van Huyck et al. conjectured, clarity of defection is able to support the efficient a_1 repeated-game equilibrium in the PD-like game S(5/6,2), but not in I(5/6,2).

5.3.3 Price Competition

Capra, Goeree, Gomez, and Holt (2002) studied a dominance-solvable game of imperfect price competition. Two firms choose prices, p_1 and p_2, from the interval [\$0.60, \$1.60]. Assume $p_1 < p_2$. Then the low-price firm earns p_1 and the high-price firm earns a fraction $\alpha \cdot p_1$ (in the experiment α is 0.2 or 0.8). If the prices are equal they each earn $(1+\alpha)p_1/2$.

The parameter α measures the degree of responsiveness of buyers to which seller named the best price. When α is close to 1, buyers still buy from the high-price seller, but at the lower price. This kind of structure reflects markets with "meet-or-release" contracts, in which buyers pledge to buy from sellers, who in turn must either meet a competitor's lower price or release the buyer to go elsewhere. When α is low, buyers gravitate toward the low-price seller. The effect of price competition is perhaps more relevant in the economy than ever before, as the Internet allows consumers to search for prices with the click of a mouse.

The imperfect price competition game is also similar to a centipede game in which players "start" at low prices. "Passing" corresponds to raising the price. Players have a joint incentive to raise prices by passing, but if a player ever expects the other player to stop passing then she wants to stop first and name the minimum price.

When $\alpha < 1$, theory predicts that Bertrand competition drives prices to the minimum of \$0.60, the unique Nash equilibrium, regardless of how far α is below 1. Capra et al. also derive quantal response equilibria, and prove that raising α leads to higher QRE prices.

Figure 5.4 shows the time series of prices in each of three sessions with $\alpha = 0.8$ (dotted lines) and three sessions with $\alpha = 0.2$ (thin lines). Averages of same-α sessions are shown with thick lines. When $\alpha = 0.2$, prices fall from around \$0.90 to \$0.70, converging reasonably close to the Nash prediction of \$0.60. But when $\alpha = 0.8$, prices start around \$1.20 and do not converge downward at all.

5.3.4 The Travelers' Dilemma

Capra et al. (1999) and Cabrera, Capra, and Gómez (2002) ran experiments on a "travelers' dilemma" which is quite similar to their price competition games (and also to p-beauty contests[14]). In the travelers' dilemma, two players simultaneously state price claims, between 80 and 200, for luggage they lost. The airline pays both players the *minimum* price. The airline also adds a reward of R to the player who stated the lower price, and subtracts a penalty of R from the player who stated the higher price.

Payoff-maximizing players will state prices one unit below what others are expected to state, in order to boost the minimum price (and hence their

[14] Nagel (1998) has run beauty contest experiments in which the player whose number is closest to p times the average earns the average, so that payoffs fall as equilibration occurs. This variant is closely related to the travelers' dilemma and the imperfect price competition game.

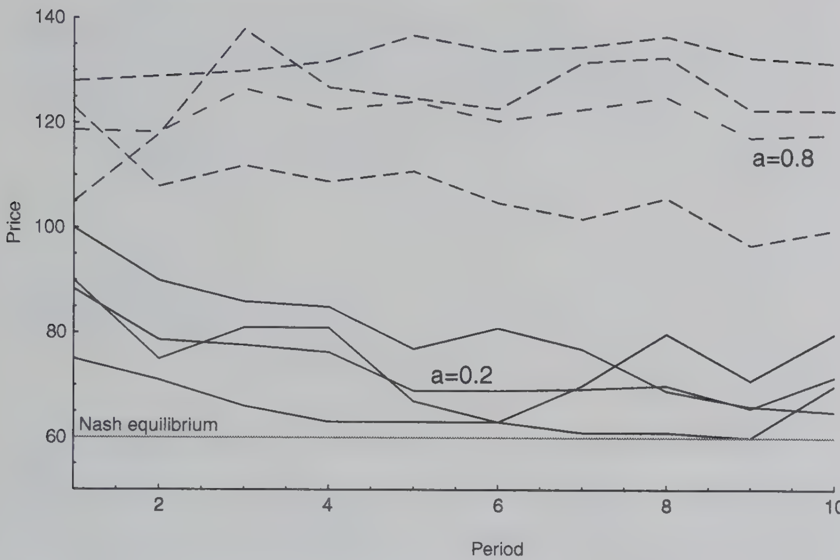


Figure 5.4. *Time series of prices with imperfect competition. Source: Based on data from Capra et al. (2002)*

payoff) while earning the reward R. The result is a "race to the bottom" in which players should choose the minimum claim, the unique Nash equilibrium. The travelers' dilemma is like a price competition game in which consumers buy from each of two sellers at the lower price but also reward and penalize sellers who stated lower and higher prices (reflecting word-of-mouth recommendations, spitefully withholding future demand from high-priced sellers, and so forth).

Figure 5.5 shows the average price claimed across the ten rounds of part A, for six values of the reward/penalty parameter R. Prices gradually converge toward the Nash equilibrium of 80 (the lowest possible price) only when $R = 50$ or 80. When R is lower (5 or 10), players actually move slightly *away* from the Nash equilibrium, toward a collusive price of 200.

Capra et al. also discuss learning theories that attempt to explain the pattern of adjustment (see Chapter 6).[15]

[15] They fit a fictitious play model which gets the direction of convergence right—it fits the upward drift when $R = 5$ and 10, and the downward drift in other conditions—but it is off by a factor of four predicting the magnitude of convergence. An experience-weighted attraction EWA does better because it is responsive to forgone payoffs and also depreciates old experience, enabling larger changes across the experiment than fictitious play (see Ho, Camerer, and Chong, 2001).

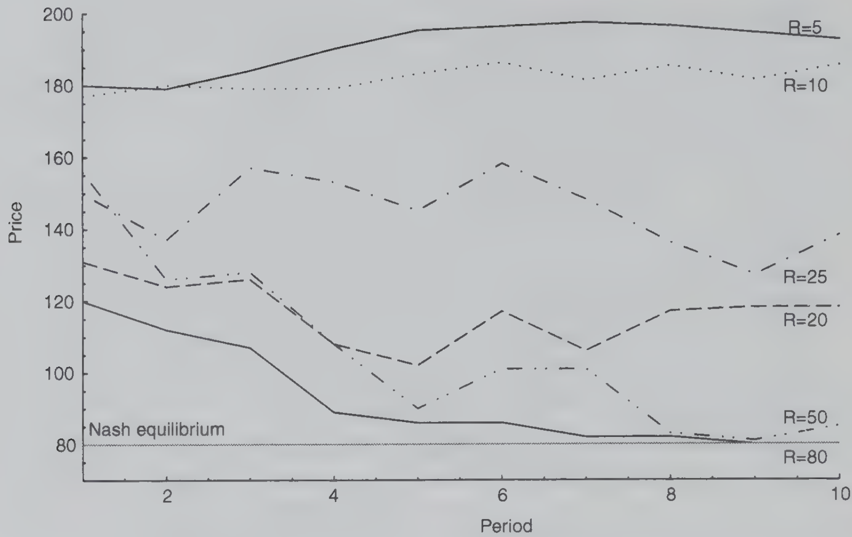


Figure 5.5. *Time series of prices in price-matching games with varying loyalty parameters R. Source: Based on data from Capra et al. (1999).*

5.3.5 The "Email Game"

The "email game" is an interesting dominance-solvable game, introduced in economics by Ariel Rubinstein (1989). Rubinstein's game is based on a well-known problem in parallel computing called the "coordinated attack problem," described by Joseph Halpern (1986). I illustrate the argument using the parameters from an unpublished experiment conducted by Barry Blecherman, David Goldstein, and myself.

In our experiments there are two states, M1 and M2, which occur with probabilities .2 and .8. Occurrence of each state triggers use of a different payoff matrix, shown in Table 5.11. The informed player 1 *always* knows the state (and hence the matrix). If the state is M1, then no messages are sent to player 2. If the state is M2, a message is sent to player 2 announcing that the state is M2. This message (and all subsequent messages) is received by the receiver player with probability .8 and intercepted with probability .2. When a message is received, the receiving player automatically sends a reply acknowledging its receipt; the reply may also be transmitted or intercepted, and so on. The message-sending process stops when a message is intercepted. There will always be a finite time at which a message gets intercepted and the process screeches to a halt.

In the coordinated attack interpretation of the game, players 1 and 2 are generals commanding divisions of an army. The enemy has troops at

Table 5.11. Payoff matrices M1 and M2 for the "email game"

		Uninformed player 2 choice	
State M1		A	B
Informed player 1 choice	A	1,1	0,−2
	B	−2,0	0,0
State M2		A	B
Informed player 1 choice	A	0,0	0,−2
	B	−2,0	1,1

location B, but may be moving some of those troops to A. The states M1 and M2 correspond to the location of the troops—if the enemy troops are moved to A, that's state M1; if they are still massed at B, that's state M2. The payoffs reflect the fact that the two generals (players 1 and 2) together can defeat the enemy if they can coordinate their attack at the right location—if they coordinate an attack at A when the state is M1, or coordinate an attack at B when the state is M2. But the enemy always has troops left behind at B, so if a single general takes troops there by himself (choosing B when the other general chooses A), he will suffer defeat and earn a negative payoff.

First notice that if the states were common knowledge, then the players would choose A in state M1 (because choosing A weakly dominates choosing B) and B in state M2 if they use payoff dominance to select the (B,B) equilibrium. If the players do not know anything about the states (other than their probabilities), there is an equilibrium in which both players choose A, another equilibrium in which both players choose B, and a mixed-strategy equilibrium. The important point is that having common knowledge of the state is essential to achieve full coordination (that is, to agree on A when the state is M1 and on B when the state is M2). Rubinstein asks: What if the players do not have common knowledge, but have "almost common knowledge," in the sense that they know others know they know . . . up to several levels of iterated knowledge? Will they play as if they have common knowledge?

In the experiment, we tried to make this tricky game as concrete as possible. Players were shown hypothetical "information sheets" listing all the messages each player sent and received. Figure 5.6 is a tree depicting sequences of messages and the associated information sheets subjects would see. For example, suppose the state is M2 (triggering the message-sending process), player 1's first message is received, player 2's acknowledgement

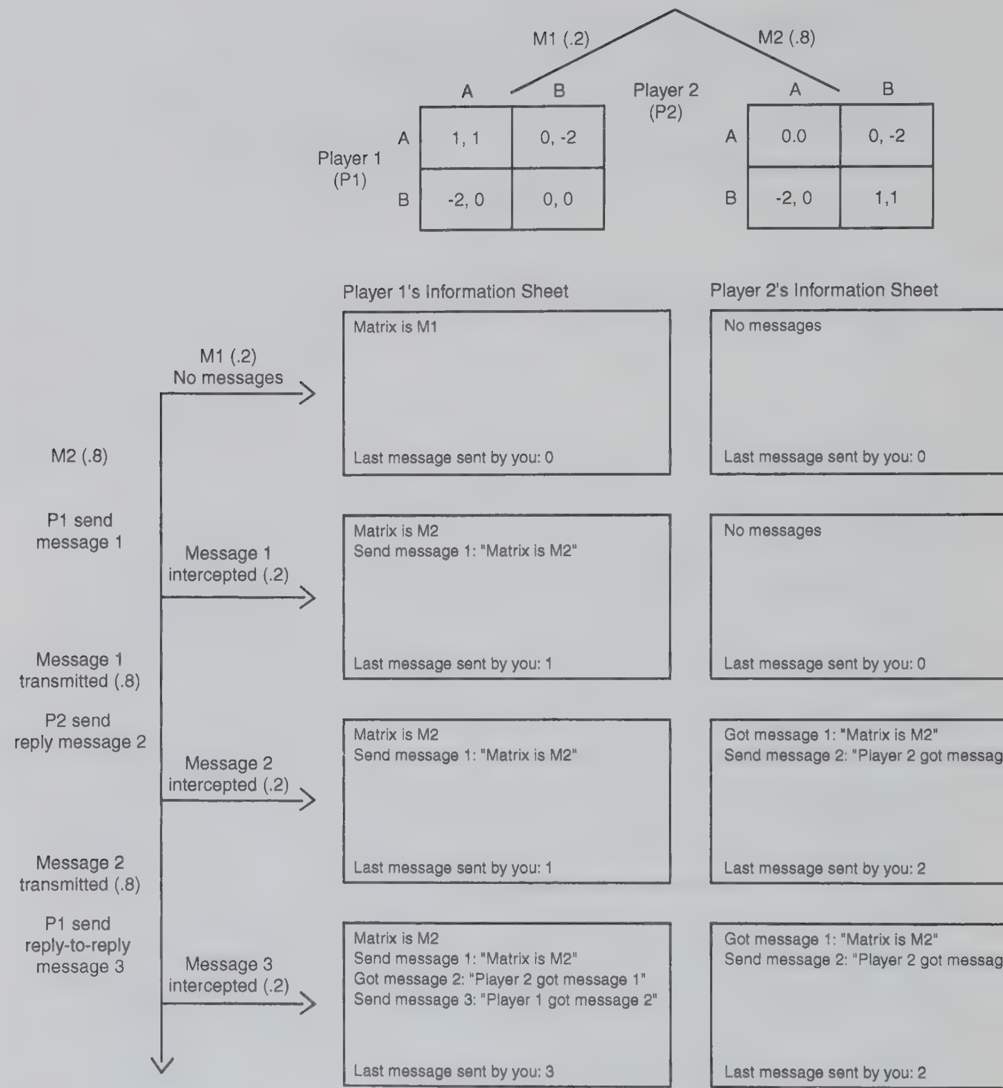


Figure 5.6. *Information tree showing information sheets of players in the "email game."*

is sent and received, but player 1's reply to player 2's acknowledgement is intercepted. Then player 2's information sheet reminds her that she received a message saying "Matrix is M2," that she sent a reply saying "Player 2 got message 1," and that she heard nothing back (the third or fourth box down on the right of Figure 5.6). The crucial point is that player 2, in this situation, cannot tell whether her message to player 1 was intercepted (third box), or whether player 1 received the message and sent a reply that was intercepted (fourth box). That is, player 1's message sheet might say "Matrix is M2, sent message" or it might say "Matrix is M2, sent message, player 2 sent reply, sent reply to reply." Since player 2 can't tell if she is in the third or fourth row of Figure 5.6, she can't be sure what player 1 knows.

However, a Bayesian probability calculation shows that the relative probabilities of being in these two situations are .56 and .44; that is, it is slightly more likely that one's message was intercepted than that the message got through and the reply was intercepted.[16]

What will actually happen if players obey iterated dominance? Start with the top row in Figure 5.6. If player 1 knows the state is M1, and obeys weak dominance, she chooses A for sure. Now consider player 2 when her information sheet says "No messages." She doesn't know whether the state is M1 (i.e., she is in the top row) or the state is M2 but player 1's message was intercepted (i.e., she is in the second row). The relative probabilities are .56 and .44. If she uses two steps of iterated dominance, she will figure out that, if she is in the top row, then player 1 will choose A. Suppose, for the sake of argument, that player 2 thinks player 1 will choose B if she is in the second row. Now if player 2 chooses A, she earns a payoff of 1 with probability .56, and a payoff of 0 with probability .44. Or she can choose B, earning a payoff of -2 with probability .56 and a payoff of 1 with probability .44. Choosing A stochastically dominates B because she has a higher chance of winning 1 (.56 instead of .44) and a lower chance of losing 2 (0 instead of .56). Therefore, player 2 should choose A if her information sheet says "No messages." It is just too risky to gamble that the state is M1 and choose B, because the state is more likely to be M1 and choosing B yields the worst payoff (-2) if player 1 obeys dominance in that state and chooses A.

Now suppose player 1 has the information sheet in rows 2–3, which says that the state is M2 and she sent a message to player 2 but no reply was received. If player 1 thinks that player 2 uses two steps of iterated dominance,

[16] The unconditional probability of getting to the third row of the information sheet is .8(.8)(.2)—that is, the state must be M2 (probability .8), the first message from player 1 must get through (probability .8), *and* the reply must get intercepted (probability .2). Similarly, the unconditional probability of being in the fourth row is .8(.8)(.8)(.2). These two probabilities are .128 and .1024. Thus, given that player 2 has the information sheet in the third or fourth rows, which happens with total probability $.128 + .1024 = .2304$, the relative probability of being in the third row is .128/.2304, or .56.

she can figure out that player 2 will choose A if her information sheet says "No messages." Suppose player 2 is likely to choose B if she has the information sheet in the third or fourth rows. Faced with her information sheet that says "Matrix is M2, send message 1" and nothing more, if player 1 chooses A then she earns 0 for sure (since she knows the matrix is M2, and choosing A always yields zero in M2). If she chooses B, she earns −2 with probability .56, and thinks she earns 1 with probability .44, an expected value of −0.68. Thus, unless player 1 is very risk seeking (or misunderstands the probability), if she thinks player 2 uses two steps of iterated dominance she will choose A.[17] The latter step in the argument is crucial. Player 1 knows for sure that the state is M2 and that choosing A guarantees a payoff of zero. But she should choose A anyway. Why? Because she knows that, if her message didn't get through, player 2 will choose A, which makes choosing B too risky.

Now consider player 2's information sheet in the third and fourth rows. (We're back where we started several paragraphs ago.) If she believes that player 1 obeys weak dominance, and that player 2 thinks she (player 1) knows this, then she will figure out that player 1 will choose A in row 3. This puts her in precisely the same decision as player 1 was, and she should choose A too—even though she knows the state is M2, and knows that player 1 knows it. The argument works by induction all the way down the tree, flipping each intuitive choice of B to a reasoned choice of A like a row of dominoes toppling one after the other. Induction leads to an unbelievable result: *No matter how many messages are passed, both players should choose A.*

Rubinstein's result is shocking because it pits a strong intuition against the steady march of inductive logic. In everyday life a couple of steps of mutual knowledge is usually enough to guarantee coordinated behavior—a message and a confirmation are all it takes. Only the most neurotic first-daters would email back and forth endlessly confirming and reconfirming in order to prevent the unraveling and mutual no-show predicted by the theory. However, the inductive logic shows that, if mismatching can be very costly to the player who chooses the nondefault option that must be talked about, then not getting a confirmation back undermines coordination, which implies that not confirming a confirmation undermines coordination, and so on. In theory, no amount of confirmation is enough![18]

[17] Notice that here we switch from iterations of dominance to iterations of rationality (Bayesian utility maximization).

[18] Game theorists have argued about whether other mathematical descriptions of common knowledge are more sensible than the one Rubinstein used (Monderer and Samet, 1989). In addition, Binmore and Samuelson (2001) showed that, if communication is voluntary (including the possibility of commonly understood refusal to reply to a message), then equilibria in which limited communications lead to coordination on B can occur. Their proof is important for explaining how people are able successfully to coordinate action in everyday life, but it is beside the point for our purposes—which is to use choices in the email game to measure the level of iterated reasoning.

Note that each step of the induction requires an additional iteration of dominance or rationality, so we can use the choices subjects make to measure indirectly how many iterations of dominance they apply.

To give the subjects enough experience, we created an unusual protocol combining the "strategy method" with a large number of randomized re-pairings. In each round, subjects stated whether they would choose A or B for each possible information sheet. Then, in each of 500 rounds, a subject was randomly re-paired with another subject (of the opposite type), realizations of the state and the message passing were simulated, and payoffs were determined by the realizations and the subjects' stated strategy choices. At the end of the 500 rounds, subjects were given a summary table of the number of times that each information sheet was generated, the number of times that their opponents chose A and B in those situations, their total payoffs for each information sheet, and their total payoff for the round. This prepackaged feedback was designed to make it as easy as possible for them to learn to play the equilibrium. Note that, although the monetary payoff in each of the 500 rounds is small, the total incentive is high because the player is essentially forced to stick to her strategy for 500 rounds. As a result, recording a B instead of an A for one information set could easily cost a subject a couple of dollars in one 500-period round, so the *marginal incentive* is very high.

So what happened? Each player's strategy in a round is a list of letters, either A or B, corresponding to their choice for each information sheet. Two sessions were conducted with the Table 5.11 payoffs, and the results are shown in the Appendix. Results from two other sessions (with the Table 5.11 payoff of -2 changed to -4) are shown in Figure 5.7. The figure shows the fraction of subjects choosing the equilibrium choice A for each number of messages (i.e., the information boxes in Figure 5.6), over fourteen sequences. In the first sequence, few subjects choose A after the first three information levels, corresponding to three levels of iterated dominance. As they learn, As are chosen for higher and higher levels of information, which means the frequency of B play dissipates. As the sequences progress, disequilibrium play shrivels up for higher and higher numbers of messages. Learning is strongly driven by losing money from choosing B for a particular information level, which precipitates switching to A about half the time.[19] When player 1s switch to A for one information level, choosing B at the next-highest level then becomes unprofitable for player 2s and a switch to

[19] With Table 5.11 payoffs, switching to A occurred eight times after a component B choice was profitable, and fifty times after B lost money. But money losing did *not* lead to switches forty-eight times. The corresponding figures for the Figure 5.7 data are nineteen switches after earnings, sixty-four switches after losing, and forty-two failures to switch. This "win–stay, lose–shift" heuristic was first reported by Messick (1967).

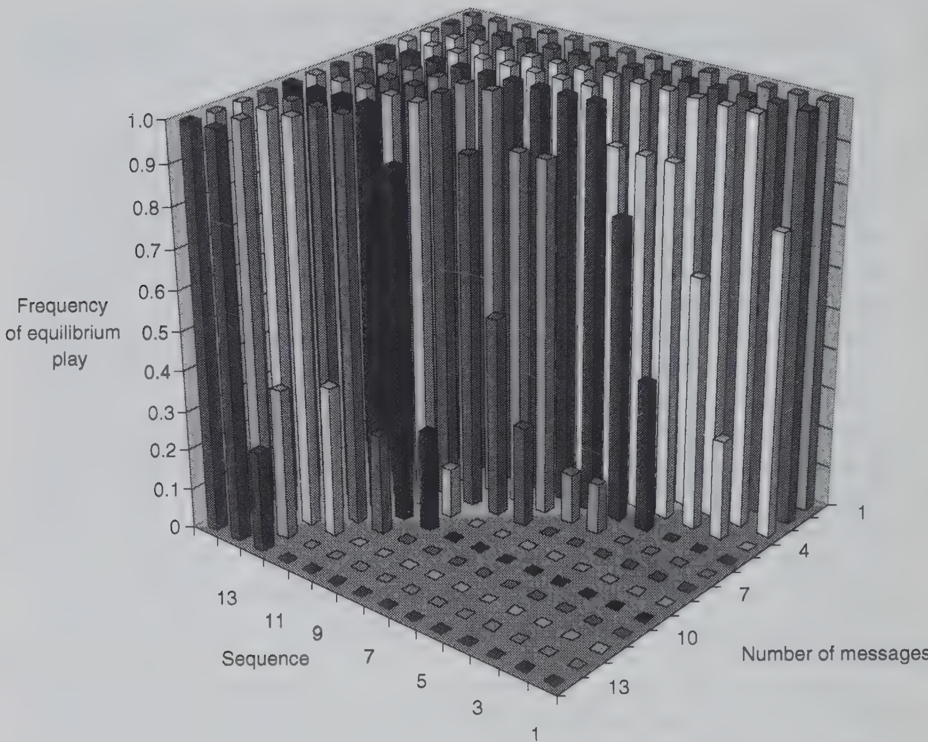


Figure 5.7. *Percentage of players choosing equilibrium given messages. Source: Camerer, Blecherman, and Goldstein (unpublished).*

A subsequently results. By sequence 13, there is *no* disequilibrium play—i.e., all subjects are playing the weird Nash equilibrium, choosing A for *all* information levels.

5.3.6 An Implementation Mechanism That Uses Iterated Dominance

Sefton and Yavaş (1996) studied iterated dominance in a neat paper on mechanism design. Consider game 1 in Table 5.12, a pure coordination game with Pareto-ranked payoffs. In pilot experiments, subjects overwhelmingly (92%) chose Red, leading to the Pareto-dominant outcome (1.20, 1.20). (Chapter 7 reports more results of this type.)

Abreu and Matsushima (1992a) consider the important problem of how a planner might implement outcomes by altering the rules of the game. For

Table 5.12. Sefton and Yavaş's game 1

	Column	
Row	Red	Blue
Red	1.20, 1.20	0,0
Blue	0,0	.60, .60

Source: Sefton and Yavaş (1996).

example, suppose you wanted to get the players to earn the Pareto-inferior (Blue, Blue) outcome in game 1. How would you do it?

Here's an example of this sort of mechanism design. Suppose the players are developers who are trying to decide at which of two locations, Redland or Blueland, they should build offices and homes. Suppose further that people like living near work. Then offices and homes are complements, in the sense that the office-builder wants the offices to be located near the homes—for simplicity, at the same location—and the home-builder wants people to be able to live near offices. Now suppose an outside agency, such as a city or state authority, decides they would rather have development take place in Blueland.[20] The agency would like to induce the developers to choose Blue(land).

The Abreu and Matsushima mechanism slices the game into T "pieces" or stages, dividing all payoffs by T; then the players play the T divided games in a sequence by specifying a series of Red or Blue plays for each of the T stages. The player who plays Red first in a sequence pays a fine of F. (Both are fined if they play Red at the same time.) In game 1, if the fine is greater than $1.20/T$ then neither player would ever want to play Red in the first of the T games, because they can earn at most $1.20/T$ in a particular stage but would then have to pay a fine of $F > 1.20/T$. If all T choices are specified at once then, by iterated deletion of dominated strategies, neither player will choose Red in the first stage, so neither will want to choose Red in the second stage (and be fined), and so on. The logical conclusion, after T steps of iterated dominance, is that no player should *ever* chose Red. Therefore, the fines "implement" the unnatural (Blue, Blue) outcome. Furthermore, in equilibrium, no player should deviate and pay the fine, so if the threat is credible and the mechanism is understood no fines will ever have to be collected (it should be "free"). In our developer example, the T stages are

[20] For example, there may be some other policy goal, such as providing employment in areas near Blueland, whose social payoffs are not reflected in the payoffs to the developers shown in the matrix.

like requiring developers to make partial investments or agreements along the way, and taxing or penalizing the first one to invest in Redland.

Notice that the Abreu and Matsushima mechanism could be easily implemented by setting $T = 1$, forcing the players to play Blue right away or pay a large fine of \$1.20. The practical value of the mechanism is clearly in allowing the planner to make F small by making T large. Having a small fine is better, because players are more likely to think that the planner really can collect the fines.

However, as Glazer and Rosenthal (1992) pointed out, as T rises, players must do more steps of iterated deletion of dominated strategies for the mechanism to work in theory. So the main virtue of the mechanism springs from an assumption that is empirically suspect. Glazer and Rosenthal (1992, p. 1438) say they would "hesitate to give long odds" on successful implementation. Abreu and Matsushima (1992b, p. 1441) responded by saying that "[our] gut instinct is that our mechanism will not fare poorly in terms of the essential feature of its construction, that is, the significant multiplicative effect of 'fines.'"

Sefton and Yavaş's subjects play T times with $F = \$0.225$, specifying a sequence of T choices of either Red or Blue. In theory, the mechanism should induce (Blue, Blue) play when $T = 8$ or 12 but not necessarily when $T = 4$. For example, when $T = 8$ the players earn \$.15 from (Red, Red) in each game piece, but are fined \$.225 if they choose Red before others do, so it always pays to choose Red one stage after the other player and avoid the fine.[21] When $T = 4$, however, the potential earnings in a period from Red is \$1.20/4=\$.30, which is greater than the \$.225 fine, so it can pay to choose Red and pay the fine anyway.

Figure 5.8 shows the proportion of all-Blue sequences for various T. In theory, there should be many all-Blue sequences for $T = 8$ or 12 and none for $T = 4$, but the opposite is true. After the first few periods, hardly any players choose all-Blue in $T = 8$ or 12, and in $T = 4$ sessions the fraction showing all-Blue rises steadily.

If subjects aren't choosing all-Blue sequences as theory predicts, what *are* they doing? After the first couple of rounds, 90% of subjects choose monotonic sequences of all-Blue followed by a series of Reds, switching to Red at some "switchpoint" round. Subjects are trying to coordinate on Red as soon as they can, to earn the higher payoff, but they try to switch from Blue to Red one period *after* the other player, to avoid the fine. In general, subjects switch in the third or fourth round (independently of T), and the switching point creeps up by about one round across the fifteen trials.

[21] (Blue,Blue) is uniquely rationalizable when $T = 8$ or 12, but when $T = 4$ there is also an equilibrium in which both players play Red all the time, happily paying the fine to earn more money in each period.

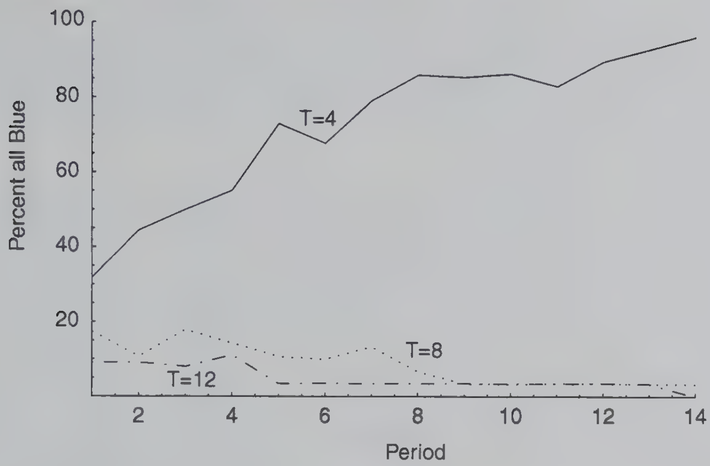


Figure 5.8. *The percentage of all-Blue sequences (Nash is all-Blue) in implementation games.* Source: *Sefton and Yavaş (1996).*

Glazer and Perry (1996) show how implementation can work in a sequential game via backward induction. Their idea is inspired by Glazer and Rubinstein (1996), who suggested that deletion of dominated strategies in a normal-form game can be made transparent as stage-by-stage comparisons in an analogous extensive-form game (see the Schotter, Weigelt, and Wilson experimental result described earlier). Katok, Sefton, and Yavaş (2002) did experiments with the sequential Glazer–Perry mechanism, but it, too, does not implement outcomes as predicted.

Katok, Sefton, and Yavaş's results are an empirical part—and the *only* empirical part—of a dialogue among theorists and experimenters. The reply by Abreu and Matsushima to Glazer and Rosenthal represents a rare, cogent written statement by two clever theorists about their perceptions of the interplay between theory and observation. First they note that the mechanism might have to be *explained* to subjects. This sort of explanation requirement should be included as part of the theory (since it is, in an everyday sense of the word, what is usually meant by "implementation"!), and suggests a good experiment. They also write: "The prospect of being at the mercy of controlled experiments is indeed intimidating" and say that "experimental results frequently defy all attempts at even approximately rational explanation" (1992b, p. 1440). The last statement is too pessimistic. In fact, this book strives to show that the opposite is true: There are hardly *any* experimental results that cannot be explained by some form of psychologically plausible theory. For example, Sefton and Yavaş's results are consistent with subjects doing three to four steps of iterated deletion of dominated strategies, guessing others will start playing Red in the third period and choosing Red one step later. Learning also causes the switchpoint to drift up over time. Both limited iterated reasoning and payoff-sensitive learning are certainly "approximately rational."

Summary: In most games reviewed in this section players exhibit two to four steps of iterated dominance in initial choices. The price competition and travelers' dilemma games are a little different because they have much larger strategy spaces (large ranges of prices). These results could be interpreted as some subjects choosing initial prices in the midpoint of a strategy space, and doing iterated reasoning in 5- or 10-point jumps. In the email game, large amounts of packaged feedback lead subjects to play an extremely counterintuitive equilibrium requiring fourteen steps of iterated deletion of dominated strategies. This result shows that the learning process can produce dominance-solvable equilibration, but perhaps only under special conditions. The Sefton–Yavaş study shows that a mechanism widely discussed in economic theory does not work as predicted because it presumes too much iterated dominance (i.e., the mechanism does not take into account "computability constraints").

5.4 When More Iteration Is Better: The "Dirty Faces" Game

A logic problem in which iterated reasoning plays a central role was first described by Littlewood (1953, pp. 3–4): "Three ladies, A, B, C, in a railway carriage all have dirty faces and are all laughing. It suddenly flashes on A: why doesn't B realize C is laughing at her? Heavens! *I* must be laughable. (Formally: If I, A, am not laughable, B will be arguing: If I, B, am not laughable, C has nothing to laugh at. Since B does not so argue, I, A, must be laughable.)" Translated from Littlewood's quaint language: If B saw A's clean face, she would infer from C's laughter that C was laughing at *her* (Ms. B), quit laughing, and start blushing (or run to the washroom to clean up). Since B doesn't blush, then A's face must be dirty too. Notice that, for A to realize this, A must think that B is rational enough to draw an inference from C, a subtle chain of reasoning. Littlewood points out that this logic can be extended link by link to an arbitrarily large number of dirty-faced ladies.

Variants of this game are widely used in conceptual discussions about iterated knowledge. The first experimental evidence is Weber (2001). In his study, players have independently drawn types, either X or O, with probabilities .8 and .2. After observing the types of *other* players—but not their own types—players take actions simultaneously in a series of rounds. At the end of each round, all the players' actions are announced. The actions

Table 5.13. Payoff table for the "dirty faces" game

		Type	
		X	O
Probability		(.8)	(.2)
Action	Up	0	0
	Down	\$1	–\$5

Source: Weber (2001).

are choices of either Up or Down. Payoffs from combinations of actions and types are shown in Table 5.13.

If players choose Up, they earn nothing. If players choose Down, they earn \$1 if they are type X and lose \$5 if they are type O. A session of the game ends when one player chooses Down. Since the prior probability of being a type X is .8, the expected payoff to choosing Down, if players know nothing about their realized type, is –\$0.20, so players will choose Up (unless they are somewhat risk preferring). Being in state X is like having a dirty face; and choosing Down is like knowing you have a dirty face and leaving to clean up.

When at least one player is a type X, the players are commonly told, "At least one player is type X." There are two cases to consider: XO case (one player is X and the other is O) and XX (both are X).

In the XO case, after the announcement, the player who is X should infer that, since at least one player is an X, *she* is the type X (this is analogous to a dirty-faced lady seeing that the *other* ladies' faces are clean). Then she should move Down. The fraction of X types who move Down right away in the XO case therefore measures the fraction of players who can make the simple rational inference that they are the X if they see an O.

The more interesting case is XX. Both players know there is at least one type X, and they know the *other* player is X, so they do not know anything about their own type. Rational players will therefore choose Up (unless they are substantially risk preferring). Then both players are told the other player's choice. Player 1, for example, is told that player 2 chose Up. If player 1 thinks player 2 is rational, she can infer that player 2 must have known player 1 was a type X. (Otherwise, by the logic in the previous paragraph, player 2 would have moved Down.) Then, in theory, player 1 can infer her *own* type from player 2's behavior—she must be an X, or player 2 would have moved Down in the first round. So player 1 should move Down in the second round if she does two steps of inference.

Table 5.14. Results for a two-player "dirty faces" game

	Trial 1		Trial 2	
	XO	XX	XO	XX
Round 1				
UU	0	7*	1	7*
DU	3*	3	4*	1
DD	0	0	0	0
Round 2 (after UU)				
UU	–	1	–	2
DU	–	5	–	2
DD	–	1*	–	3*
Other (UD)	–	–	1	–

Source: Weber (2001).

Weber ran two-player and three-player games. The three-player results are similar so I'll report only the simpler two-player results, which are shown in Table 5.14. The predicted patterns of behavior are marked by an asterisk. Cumbersome information passing enabled Weber to do only two trials, so there is limited time for learning to occur.

In the XO condition, players pass the basic rationality test seven out of eight times by choosing Down when they are type X. In the trickier XX condition, players are predicted to choose UU in the first round, followed by DD. There *is* a lot of UU choice in the first round, but in most second rounds only one of two type Xs chooses D (i.e., DU in the table, rather than both choosing D as predicted by rationality). Counting the number of individual subjects who are behaving rationally (rather than pairs), 87% are rational in the XO condition, and 53% in the second round of the XX condition.[22] Weber also shows that three statistical variations on Nash equilibrium do a fairly good job of explaining the data.[23]

Weber's results are important because Caltech students are selected for their skill at logic puzzles such as the dirty faces game (and they know other

[22] That is, fifteen out of twenty-eight XX subjects choose D in round 2 after observing UU in round 1, pooling across trials.

[23] The three models are: a "noisy Nash" model, in which players play Nash with error, choosing correctly about 75% of the time; a probabilistic information model, which adds the assumption that players ignore information revealed by choices of others in earlier rounds with a certain probability (estimated to be around .60); and a quantal response equilibrium. All three are substantial improvements over a random choice model, and show that there is hope to modify standard concepts minimally to make sense of departures from Nash equilibrium.

students are very skilled). The extent of their iterated reasoning—half of them do not do two steps—is a plausible upper bound on iterated reasoning by most people in abstract games such as these.

Second, in the price competition, email, centipede, and mechanism games, players always do better if others have limited rationality. The dirty faces game is the opposite. If players have limited rationality they will choose Down when they should choose Up (and vice versa), harming the other players' ability to draw inferences about their own types. The fact that iterated reasoning is quite limited in the dirty faces game supports the conclusion that choices in the other games reveal limited reasoning (rather than pure cooperativeness).

5.5 The "Groucho Marx" Theorem in Zero-Sum Betting

Sovik (1999) studied a zero-sum betting game (replicating Sonsino et al., 1998). In the game there are four states, A–D, which are equally likely. Players have (commonly known) information partitions which tell them the set of possible states that may have occurred, for a particular state realization. After learning their information, players choose whether to bet or not. Betting is zero-sum, so that one player wins what another loses if both bet.

Table 5.15 shows the information partitions and payoffs for a type I and type II player. If the state is A, for example, player I knows only it could be A *or* B, but type II knows it is definitely A. The type II player should certainly bet in state D (since she knows the state is D and will win 16 if the other player bets also). Type II should certainly *not* bet in state A (since she will lose 32 for sure). If the type I player thinks type II is rational, then she knows II will never bet when the state is A, so for type I to bet when her information is the partition (A,B) means she can never win. If type II knows that type I

Table 5.15. *Information partitions in betting game*

	State			
	A	B	C	D
Type I	(A,B)		(C,D)	
I payoff	32	−28	20	−16
Type II	A	(B,C)		D
II payoff	−32	28	−20	16

Source: Sovik (1999).

is rational (and thinks, I think she, type II, is rational) then she infers I will never bet when her information is (A,B), so that betting when her own type II information is (B,C) will result in the loss of 20 (when the state is C). Iterating one more step, if type I knows that II is rational, and knows that II thinks I thinks II is rational, then she can infer that II will never bet in (B,C), so that betting in (C,D) is a lost cause for type I. Therefore, no bets will ever take place (in theory).

This result is an example of the "Groucho Marx theorem" (Milgrom and Stokey, 1982): Players who are trading only for the purpose of speculation (rather than to shift risk or to enhance the excitement of watching an event unfold), should never trade, because in zero-sum betting it never pays to bet with somebody who is willing to bet with you. (The theorem was named by Milt Harris, after the famous quip by Groucho Marx that he would never join a club that would have him as a member.) A similar intuition is present in the winner's curse in uninformed bargaining and common value auctions (namely, take somebody's willingness to trade, or to let you outbid them, as a signal—which is bad for you—about their information). But in those situations there is usually *some* trade, in equilibrium, because correcting for anticipated adverse selection still leaves some (expected) gains from trade, but in zero-sum games there are no gains and hence there should be no trade.

The Groucho Marx no-trade result is quite unintuitive, since it requires many levels of iterated reasoning. The no-trade prediction appears to be violated dramatically, in the sense that there is a huge volume of zero-sum betting in options, futures, and commodity markets—a trillion dollars worth of foreign exchange trading every day in New York—which is too high to be due to risk hedging (e.g., Sen, 2002). Experiments provide a nice clean way to see whether the basic logic underlying the no-trade result is transparent to subjects, and if not, how quickly it can be learned.

Sovik's design was motivated by an interest in the robustness of pioneering experiments by Sonsino et al. (1999). In their experiments, not betting gave a sure payoff of one Israeli shekel; betting gave the outcomes in Table 5.15 if both sides chose to bet. Sonsino et al. conducted several different designs. The minimum and maximum betting rates—the fraction of trials in which subjects chose to bet—in the first 50 periods (out of 250) and the last 50 periods are shown in Table 5.16, for each partition. Two odd regularities stick out. Type II players should never bet in state A (they know they will lose 32 for sure, rather than earning 1); yet they bet 6–24% of the time in the first 50 periods and 4–20% of the time in the last 50 periods! And they should *always* choose to bet in state D, since they cannot lose and can win if the type I player bets. But they choose to bet only 78–95% of the time in state D the first 50 trials, and 81–100% in the last 50 trials. These rates of betting at a certain loss, and failing to bet for a certain gain, are not huge, but mistake rates of this magnitude are unusual. More importantly, if

Table 5.16. *Bet frequencies in a zero-sum betting game*

	State			
	A	B	C	D
Type I betting rates	(A,B)		(C,D)	
Sonsino et al.				
First 50 (min,max)	0.31–0.70		0.49–0.63	
Last 50 (min,max)	0.12–0.48		0.37–0.65	
Sovik				
First 12	0.44		0.50	
Sovik				
Last 12	0.11		0.60	
Type II betting rates	A	(B,C)		D
Sonsino et al.				
First 50 (min,max)	0.06–0.24	0.61–.073		0.78–0.95
Last 50 (min,max)	0.04–0.20	0.58–0.70		0.81–1.00
Sovik				
First 12	0.00	0.69		0.99
Last 12	0.00	0.55		1.00

Source: Sovik (1999).

the type IIs are betting when they should not, their aberrant behavior slows down the rate at which type Is learn not to bet when their partition is (A,B), which slows down type IIs learning not to bet when they know (B,C), and so on. As a result, Sonsino et al. see very slow convergence toward the no-trade equilibrium even over 250 trials.

Sovik altered the Sonsino et al. design in several ways: paying money rather than binary lottery points; raising the stakes (Table 5.15 points were worth 2.5 pesetas each, about \$0.02, eight times higher than in Sonsino et al.); playing in random rematches as well as fixed pairings; and using only 24 periods rather than 250. In the first 12 periods players were given a specific information set and asked whether they wanted to bet. In the last 12 periods Sovik used the "strategy method," asking subjects whether they would bet for each possible partition (then one state was drawn and their full strategy was used to determine what they did do).

The results are summarized and contrasted with those of Sonsino et al. in Table 5.16. First note that type II players *never* bet when the state was A, and *almost always* bet when the state was D. Despite these lower rates of dominance violation, iterated reasoning was still quite limited because type

Is bet almost half the time when their information was (A,B) in the first 12 trials. However, this rate dropped to 11% in the last 12 trials.

The strategy method shows full information-contingent strategies. One-third of the type Is never bet, and half chose not to bet in (A,B) but bet in (C,D) (which is consistent with two levels of iterated reasoning). Type IIs chose to bet in (B,C) and in D half the time, which is inconsistent with three steps of iterated reasoning, and they bet only in D half of the time. Overall, the evidence is consistent with two steps of iterated reasoning by most players.

5.6 Structural Models of Decision Rules and Levels of Reasoning

This chapter has highlighted how games can be used to measure the frequency of violation of different numbers of steps of deletion of dominated strategies. A problem with this method is that subjects who appear to obey dominance may be doing so even though they use specific decision rules that do not *generally* respect dominance. For example, in beauty contest games, student players sometimes choose numbers such as 21, which correspond to two levels of iterated reasoning (from a starting point of 50). But they may just be choosing their age in years, rather than truly thinking iteratedly. When they turn 67 they'll start violating dominance!

The papers described in this section give a clear answer by positing decision rules directly and seeing how well they explain what players do.

Stahl and Wilson (1995) proposed a model in which players have different degrees of strategic sophistication. The model was foreshadowed by Stahl (1993) and Nagel (1995) (and also Holt, 1999, which first circulated in 1993).

They distinguished three "levels of reasoning": Level-0 players choose uniformly; and level-k players choose best responses to level-$(k-1)$ choices, for $k \in \{1, 2\}$. Obviously, one could extend the number of levels arbitrarily, but Stahl and Wilson stopped at two, and introduced some other rules as well.

Some notation is necessary to describe precisely what they did and appreciate its compact parsimony. Consider symmetric games with two players and three strategies. Let $j \in \{1, 2, 3\}$ denote strategies, and U denote a payoff matrix for game I. U_{jk} denotes the payoff to a row player who chooses j when column chooses k. Vectors p will denote beliefs about the frequency of choices by others. P_0 is the uniformly distributed belief over the three choices. The expected payoff from strategy j given belief p_i is therefore the matrix product $U_j p_i$. Denote a best response to belief p by $b(p)$. Denote the Nash equilibrium choice frequency vector by p_j^{Nash}.

First define a general family of beliefs q_j which reflect the assumption that ϵ of the population play a noisy best response to P_0, and the remaining $1-\epsilon$ play Nash. Formally,

$$q_j(\mu, \epsilon) = \epsilon \cdot e^{\mu U_j P_0} / \sum_k e^{\mu U_k P_0} + (1-\epsilon) p_j^{\text{Nash}}, \tag{5.6.1}$$

where μ is a response sensitivity parameter, expected payoffs from the beliefs q_j are denoted $y_j(\mu, \epsilon) = U_j q_j(\mu, \epsilon)$. Players are assumed to choose according to expected payoffs with noise parameter γ, according to $P_j(\gamma, \mu, \epsilon) = e^{\gamma y_j(\mu,\epsilon)} / \sum_k e^{\gamma y_k(\mu,\epsilon)}$.

This simple two-step structure includes an amazing variety of decision rules corresponding to different values of the parameters μ (perceived noise in others' choices), γ (noise in a player's own choices), and ϵ (the fraction of perceived non-Nash players). Because each subject makes only twelve choices, it is hard to reliably estimate individual parameters separately, or even search for clusters. Instead, Stahl and Wilson restricted players to belong to a small set of possible types (specific parameters or ranges) and estimated a model in which the population is a mixture of those types. They posited five types,[24] with relative frequencies α_0 to α_4:

1. Level-0 types just choose each strategy equally often, so $\gamma = 0$.
2. Level-1 types assume that all others are level-0 types, so $\mu = 0, \epsilon = 1$ (i.e., all others choose randomly), and $\gamma > 0$.
3. Level-2 types assume that all others are level-1 types, so $\epsilon = 1$ and $\mu > 0$ (their beliefs about what level 1s will do are generated by noisy best responses to the belief P_0), and $\gamma > 0$.
4. Naive Nash types think others will play Nash, so $\epsilon = 0$ and $\gamma > 0$.
5. Worldly Nash types think some others will play Nash, but some players compute expected payoffs from P_0 beliefs or best responses $b(P_0)$, so $0 < \epsilon < 1$ and $\gamma > 0$.

Subjects play twelve symmetric 3×3 games, shown in Table 5.17. The games were selected to have various strategic properties.[25] The table shows payoffs to the row player and lists which decision rules select which strategies. The actual choice frequencies are also shown in the table.

[24] They actually include a sixth "rational expectations" (RE) type, which is assumed to know the relative frequencies of all five types and their own type. These types are estimated to be quite infrequent and are dropped from most analyses.

[25] Games 1, 5, and 12 are strict dominance solvable; games 3 and 9 are weak dominance solvable; games 4, 7, and 11 have unique mixed-strategy Nash equilibria; and games 2, 6, 8, and 10 have pure-strategy Nash equilibria (in which the Nash, $b(P_0)$, and $b(b(P_0))$ strategies are distinct). When Nash is listed more than once in the table, the probability with which each strategy is used in the mixed-strategy equilibrium is given in parentheses.

Table 5.17. *Stahl and Wilson's payoff tables, model predictions, and actual frequencies*

		Payoff table			Actual frequency	Mixture model prediction	Decision rules selecting strategy
Game 1		T	M	B			
	T	25	30	100	0.15	0.19	$b(P_0)$
	M	40	45	65	0.83	0.76	Nash, $b(b(P_0))$,RE
	B	31	0	40	0.02	0.06	Dominated
Game 2		T	M	B			
	T	75	40	45	0.63	0.51	Nash
	M	70	15	100	0.25	0.28	$b(P_0)$,RE
	B	70	60	0	0.13	0.21	$b(b(P_0))$
Game 3		T	M	B			
	T	75	0	45	0.10	0.06	Dominated
	M	80	35	45	0.33	0.41	Nash,$b(b(P_0))$,RE
	B	100	35	41	0.56	0.53	$b(P_0)$
Game 4		T	M	B			
	T	30	50	100	0.54	0.56	Nash (.67),$b(P_0)$,RE
	M	40	45	10	0.31	0.25	Nash (.27),$b(b(P_0))$
	B	35	60	0	0.15	0.19	Nash (.06)
Game 5		T	M	B			
	T	10	100	40	0.29	0.22	$b(P_0)$
	M	0	70	50	0.06	0.10	Dominated
	B	20	50	60	0.65	0.68	Nash,$b(b(P_0))$,RE
Game 6		T	M	B			
	T	25	30	100	0.23	0.27	$b(P_0)$,RE
	M	60	31	51	0.42	0.43	Nash
	B	95	30	0	0.35	0.30	$b(b(P_0))$

Table 5.17. *(continued)*

		Payoff table			Actual frequency	Mixture model prediction	Decision rules selecting strategy
Game 7		T	M	B			
	T	30	100	50	0.44	0.45	Nash (.53), $b(P_0)$,RE
	M	40	0	90	0.35	0.25	Nash (.17)
	B	50	75	29	0.21	0.30	Nash (.30), $b(b(P_0))$
Game 8		T	M	B			
	T	0	60	50	0.25	0.26	$b(b(P_0))$
	M	100	20	50	0.25	0.31	$b(P_0)$,RE
	B	50	40	52	0.50	0.43	Nash
Game 9		T	M	B			
	T	40	100	65	0.54	0.36	$b(P_0)$
	M	33	25	65	0.02	0.12	Dominated
	B	80	0	65	0.44	0.52	Nash, $b(b(P_0))$,RE
Game 10		T	M	B			
	T	45	50	21	0.81	0.50	Nash
	M	41	0	40	0.06	0.22	$b(b(P_0))$
	B	40	100	0	0.13	0.28	$b(P_0)$,RE
Game 11		T	M	B			
	T	30	100	22	0.27	0.42	Nash (.37), $b(P_0)$,RE
	M	35	0	45	0.08	0.19	Nash (.13)
	B	51	50	20	0.65	0.40	Nash (.50), $b(b(P_0))$
Game 12		T	M	B			
	T	40	15	70	0.54	0.63	Nash, $b(b(P_0))$,RE
	M	22	80	0	0.06	0.06	Dominated
	B	30	100	55	0.40	0.31	$b(P_0)$

Source: Stahl and Wilson (1995).

Table 5.18. Stahl and Wilson's parameter estimates for mixture models

Parameter	Interpretation	Estimate	Standard error
γ_1	Level-1 response sensitivity	0.218	0.043
μ_2	Level-2 perceived γ_1	0.461	0.062
γ_2	Level-2 response sensitivity	3.079	0.574
γ_3	Nash-type response sensitivity	4.993	0.936
μ_4	Worldly Nash response sensitivity of non-Nash types	0.062	0.006
ϵ_4	Worldly-type perception of percentage of non-Nash types	0.441	0.077
γ_4	Wordly-type response sensitivity	0.333	0.055
α_0	Frequency of level-0 types	0.175	0.059
α_1	Frequency of level-1 types	0.207	0.058
α_2	Frequency of level-2 types	0.021	0.020
α_3	Frequency of Nash types	0.167	0.060
α_4	Frequency of worldly Nash types	0.431	0.078
Log-likelihood	−442.727		

Source: Stahl and Wilson (1995).

Maximum-likelihood parameter estimates are shown in Table 5.18. There appear to be hardly any level-2 types (2%), roughly equal numbers of level-0, level-1, and naive Nash types, and 43% worldly Nash types. Thorough further analyses show that estimates are robust to cross-game forecasting and classify most subjects reliably.[26] All these frequencies are precisely estimated.

The expected payoffs for different types given the actual data range from 43.0 for the level-0 types to 46.6 for naive and worldly Nash types. These payoffs are not significantly different, which is a reminder that payoff losses or selection pressures against boundedly rational (low-level) types may be weak (Stahl, 1993).

Costa-Gomes, Crawford, and Broseta (2001) investigated decision rules and iterated dominance using a design similar to Stahl and Wilson's, but which also measures how subjects attend to payoff information (as in Camerer et al., 1994, and Johnson et al., 2002; see Chapter 4). Subjects play eighteen two-player normal-form games with two, three, or four strate-

[26] Stahl and Wilson also tested for the presence of a "sophisticated" or "perfect foresight" type who chooses best responses to the actual relative frequencies observed in the experiment, but "soundly rejected" that hypothesis.

gies, with no feedback (deliberately to disable learning). The games were constructed to enable identification of several decision rules, including different levels of iterated dominance. The nonstrategic decision rules are: naive or level 1 (choose strategies with the highest average payoff, averaging payoffs equally); optimistic or maximax; pessimistic or maximin; and altruistic (maximize the sum of the two players' payoffs). There are also five strategic types (which, by definition, think about what other players might do): L2, which best-responds to naive; D1, which does one round of deleting dominated decisions, then best-responds to a uniform prior on the remaining decisions; D2, which does two rounds of deletion then best-responds to remaining strategies; sophisticated, who guess accurately the proportion of the population that chooses each nonequilbrium strategy; and (Nash) equilibrium. By design, different rules do not overlap much in the strategies they choose.[27] These games are therefore an unusually tough test for equilibrium, compared with most experiments, in which predictions of equilibrium and other decision rules overlap.

Subjects participated in one of three conditions: Baseline (B); Open Boxes (OB); and Training (TS). In Baseline (B), payoffs were hidden behind boxes that could be opened and closed only by clicking a mouse. As discussed in Chapter 4, this information display enables the experimenter to "get behind the subjects' eyes" and see what subjects see. If decision rules are treated as algorithms that demand certain kinds of information, knowing what information subjects use helps identify the rules. In the OB treatment, the boxes were always open; comparing this with the Baseline tests whether having payoffs in boxes affects what strategies subjects choose. In the training treatment, TS, subjects were rewarded only for choosing equilibrium strategies (after receiving general instruction in game-theoretic reasoning). Their behavior established an empirical baseline for how subjects who *were* reasoning game-theoretically looked up information, which could then be compared with lookups by B subjects. The results can be summarized in three parts: consistency of strategies with iterated dominance; estimates of decision rules used by subjects; and information search patterns. Table 5.19 shows the games Costa-Gomes et al. used and frequencies with which different strategies were chosen.

The fractions of B and OB subjects choosing equilibria are about 90%, 65%, and 15% when equilibria require 1, 2, and 3 levels of iterated dominance, corroborating many other results reported in this chapter. However, the trained subjects (TS conditions) are able to execute equilibrium reasoning almost perfectly, 90–100% of the time, including three levels of iterated

[27] In the original version of their paper, the L2, D1, and D2 types were not present and a large fraction of players were classified as sophisticated.

Table 5.19. *Games and actual frequencies of Costa-Gomes et al.*

Game	Row choice	Column choice L	Column choice R	Actual frequency
2b	U	38,57	94,23	0.92
	D	14,18	45,89	0.08
Actual frequency		0.72	0.28	
3a	U	75,51	42,27	0.70
	D	48,80	89,68	0.30
Actual frequency		0.92	0.08	
3b	U	55,36	16,12	0.72
	D	21,92	87,43	0.28
Actual frequency		0.94	0.06	
4b	T	68,46	31,32	0.41
	M	47,61	72,43	0.06
	B	43,84	91,65	0.53
Actual frequency		0.89	0.11	
4c	T	51,69	82,45	0.92
	M	28,37	57,58	0.00
	B	22,36	60,84	0.08
Actual frequency		0.56	0.44	
5b	T	74,62	43,40	0.14
	M	25,12	76,93	0.11
	B	59,37	94,16	0.75
Actual frequency		0.70	0.30	
8b	T	71, 49	28, 24	0.22
	M	46, 16	57, 88	0.08
	B	42, 82	84, 60	0.70
Actual frequency		0.47	0.53	
9a	T	45,66	82,31	0.92
	TM	22,14	57,55	0.00
	BM	30,42	28,37	0.00
	B	15,60	61,88	0.08
Actual frequency		0.64	0.36	

Table 5.19. *(continued)*

		Column choice				
		L	M	R		
4a	T	70,52	38,29	37,23	0.70	
	B	46,83	59,58	85,61	0.30	
Actual frequency		0.86	0.0	0.14		
4d	T	42,64	57,43	80,39	0.89	
	B	28,27	39,68	61,87	0.11	
Actual frequency		0.78	0.03	0.19		
6b	T	64,76	14,27	39,61	0.61	
	B	42,45	95,78	18,96	0.39	
Actual frequency		0.17	0.11	0.72		
7b	T	56,78	23,53	89,49	0.44	
	B	31,35	95,64	67,91	0.56	
Actual frequency		0.19	0.06	0.75		
		Column choice				
		L	ML	MR	R	
9b	T	67,46	15,23	43,31	61,16	0.83
	B	32,86	56,58	38,29	89,62	0.17
Actual frequency		0.86	0	0	0.14	

Source: Costa-Gomes, Crawford, and Broseta (2001).

dominance. This excellent performance establishes that game-theoretic reasoning is *not* computationally difficult, per se, but is simply *unnatural* for most subjects (as seen in Johnson et al., 2002). Costa-Gomes et al. joke, only half-facetiously, that after an hour of training the subjects apply game theory better than his undergraduates do after a semester! Thus, strategic thinking seems to be more like learning to windsurf, ski, or fly an airplane, activities that require people to learn skills which are unnatural but teachable, and less like weight-lifting or dunking a basketball, where performance is constrained by physical limits.

Since different decision rules point to different strategies in each game, by looking at a player's strategy choices the player can be classified according to which decision rule she uses most often. Costa-Gomes et al. did this by assuming that each player uses one decision rule, but trembles or errs.

Table 5.20. Estimated frequency of decision rule types

Decision rule	Expected payoff ($)	Information used	
		Decisions only	Decisions + search
Altruistic	17.11	0.089	0.022
Pessimistic	20.93	0.000	0.045
Naive	21.38	0.227	0.448
Optimistic	21.38	0.000	0.022
L2	24.87	0.442	0.441
D1	24.13	0.195	0.000
D2	23.95	0.000	0.000
Equilibrium	24.19	0.052	0.000
Sophisticated	24.93	0.000	0.022

Source: Costa-Gomes, Crawford, and Broseta (2001).

Subjects were then classified by finding a rule, and error rate, that maximized the likelihood of observing the pattern of choices subjects actually made. Similar procedures were used by Harless and Camerer (1994, 1995) and El-Gamal and Grether (1995).

Table 5.20 shows the fraction of subjects estimated to be of each type, using either the decisions or decisions plus information search data, and the expected dollar payoff of those rules. Using decisions alone, about 20% of the subjects are classified in each of the naive or D1 categories, and 44% are classified as L2. Adding information search data sharpens this classification substantially—then 45% are classified as naive and *none* are classified as D1. Note also that the expected payoffs to the reasoning-level, equilibrium, and sophisticated types are all around $24 for the session. Although the sophisticated types' expected earnings are highest (by definition), the L2 types earn only $0.06 less.

Information search measures can be analyzed in various ways. The basic information measures are: the number of times the boxes are opened ("lookups"), the length of time boxes are open ("gaze times"), and transitions from box to box.

The trickiest part of the project of Costa-Gomes et al. was creating plausible restrictions on how the information search measures vary with players' decision rules. This requires one to take the decision rules very seriously, and ask how information-processing measures differ across rules. Big differences come from whether players open own-payoff boxes differently than other-payoff boxes. In their display, like the usual normal-form game matrix,

Table 5.21. Predicted and actual payoff-box transitions (percent)

Subject/decision rule	Up–down, own-payoff		Left–right, other-payoff	
	Predicted	Actual	Predicted	Actual
TS (equilibrium)	>31	63.3	>31	69.3
Equilibrium	>31	21.5	>31	79.0
Naive/optimistic	<31	21.1	—	48.3
Altruistic	<31	21.1	—	60.0
L2	>31	39.4	=31	30.3
D1	>31	28.3	>31	61.7

Source: Costa-Gomes, Crawford, and Broseta (2001).
Note: Pessimistic, D2, and sophisticated not classified.

players who are computing their own expected payoffs will shift *up and down* within an own-payoff column; that is, they will compare the payoffs with different row strategies (fixing a column, which corresponds to a fixed choice by the other player). Note that, in their display, the normal-form matrix is divided in half, with one's own payoffs in one half and the other's payoffs in the other half. A player who is naive, optimistic, or altruistic will make *left-to-right* transitions, comparing across columns to compute a strategy's average, minimum, or maximum payoff. Different rules also allow different transitions among other-payoff cells. For example, equilibrium players will find best responses for *other* players by making left-to-right transitions among other-payoff cells. Subtle calculations predict that equilibrium players will make up–down transitions among own-payoff cells more than 31% of the time (i.e., more than 31% of the transitions will be up–down rather than right–left). Other types will make those transitions less than 31% of the time. Therefore, the frequency of those transitions is one way of testing whether the classifications of subjects by their decisions corresponds to their information-processing behavior.

Table 5.21 shows these predictions and the actual transition frequencies for players who were classified into each type by the earlier analysis on their choices. TS subjects make equilibrium-type own-payoff transitions about two-thirds of the time; other types make those transitions less often. Equilibrium subjects also make many more left–right transitions in other-payoff cells than sophisticated subjects do, though the opposite holds for own-payoff transitions. Keep in mind that the row classification in Table 5.21 was done *purely* by using decisions. The positivist idea that decision rules are simply predictions about choices, rather than necessarily constraining details of a thinking process, allows the possibility that the decisions a player makes are uncorrelated with information-processing measures. But decisions and search patterns *are* correlated. Put the other way around, if the classification had been done by information search first, and choices were predicted from those search patterns, the cognitive data would have helped predict what subjects would actually choose.

Information measures can also be used to classify individuals. Costa-Gomes et al. explained the lookups necessary to execute each decision rule and defined several levels of compliance for these lookups. The weakest level of compliance ("occurrence") requires a subject to have *all* the necessary lookups *somewhere* in a trial's entire lookup sequence. The strong requirement ("adjacency") is that payoffs that must be compared to execute a decision rule must occur next to each other in a sequence.[28] For example, an altruistic subject is assumed to add together own- and other-payoffs in a particular cell. Occurrence states that she has both of the payoffs from a particular cell *somewhere* in her lookup sequence. Adjacency states that the payoffs must be next to each other in the looked-up sequence. Compliance with these requirements is then treated as a discrete variable that can be used to classify subjects simultaneously by strategy choice and lookup sequence compliance.

Since adjacency is a very strong requirement, subjects can be usefully classified by whether adjacency is high (occurring on 67–100% of the trials) or low (0–33%), assuming that occurrence is 100%. Cross-classifications by decisions and high and low adjacency, assuming perfect compliance with occurrence, are summarized in Table 5.22. The table shows rates of high and low compliance with predicted adjacency patterns (columns) for subjects classified into various rules (rows). For example, subjects classified as L2 by their decisions had 85% strong compliance and 3% weak compliance with adjacency lookup restrictions. A tight match between predicted lookups and resulting decisions would yield compliance rates on the diagonal (ignoring the TS row) with H (L) figures that are higher (lower) than other rates in the same column. For example, the 85% high compliance by L2 decision types with L2 lookup patterns is greater than the rate of high compliance in the same column for all other types (which range from 42 percent to 64 percent for equilibrium players) and the 3% low compliance rate is lower than all except equilibrium. In most cases the classification is not bad. For example, within-column ranks of the diagonal figures for high compliance provide answers to the following question: Judging from high compliance with type x lookups, how do players classified as type x by decisions rank (compared with the other four classified types)? Across columns, these ranks are 1, 2,

[28] These requirements could be considered boundary conditions on what subjects could do if they have enough working memory (weak requirement) or what they must do if they have very little working memory (strong requirement).

Table 5.22. *Decision rule classification by strategy and information lookup compliance*

Choice-classified decision rule (percent)	High, Low compliance with decision rule (percent)								
	Altruistic	Pessimistic	Naive	Optimistic	L2	D1	D2	Equilibrium	Sophisticated
TS	3, 50	44, 36	83, 0	86, 14	76, 0	92,1	92, 1	*96, 1*	75,1
Altruistic (4)	**78,11**	56,33	53,42	97,3	47,39	36,56	33,56	31,56	28,56
Naive/optimistic (24)	9,53	85,9	*89,3*	*96,4*	42,3	45,20	43,23	26,28	23,27
L2 (51)	8,58	72,9	78,0	80,20	**85,*3***	57,9	54,10	49,12	46,12
D1 (15)	23,26	59,16	63,6	77,23	53,6	*48,14*	45,15	42,17	38,21
Equilibrium (4)	6,86	100,0	97,0	100,0	64,0	69,14	67,14	**56**,*19*	53,28

Source: Costa-Gomes, Crawford, and Broseta (2001).

Note: Pessimistic, D2, and sophisticated types were never classified by decision rule. *Italics* denotes cells that should have highest, H, and lowest, L, compliance in their column; **boldface** denotes cells that should and *do* have the best compliance in their columns.

3, 1, 3, 1 (random ranks would average 3). The classification is not perfect, but it is not bad. An analysis of posterior probabilities of each subject being of a particular type shows that 80% of the subjects have a posterior on one type of .90 or more.

Summary: Studies of decision rules find a wide variety of rules which explain aggregate decisions rather well. In Stahl and Wilson's classification, around 20% of subjects are level 0 (choosing randomly) or level 1 (best-responding to level 0). Another half of the subjects are "worldly Nash" who think some subjects will choose equilibrium strategies and others will make level-0 or level-1 choices, and respond accordingly. In Costa-Gomes, Crawford, and Broseta's study, large percentages appear to be naive or optimistic, and about half seem to believe that others are best-responding to random choice (level 2). They also show how measuring the information subjects gather is crucial to help figure out what their decision rules are. For example, from decisions alone about 20% of subjects appear to choose according to naive or D1 rules. But search data adjust these classifications to 45% and 0 percent. A study that used only choice data would have drawn the wrong conclusion. They also note that about 10, 35, and 85% of untrained subjects violated 1, 2, and 3 steps of iterated dominance, which is comparable with results seen throughout this chapter.

5.7 Theories

As simple bargaining and contribution games have inspired a flurry of theorizing (see Chapter 2), the regularity in limited thinking in dominance-solvable games has led immediately to different theories which are precisely specified and ready to apply to more games.

5.7.1 Multiple Types

One approach is to classify players into types, based on the number of steps of reasoning they appear to do. Stahl (1993) deserves much credit for this idea (as does Debra Holt, whose 1999 paper first appeared in 1993 and used types to analyze Cooper et al. coordination data, discussed in Chapter 7). In Stahl and Wilson's (1995) approach, the types included iterated steps of thinking and other types (such as equilibrium and "worldly" players). Costa-Gomes, Crawford, and Broseta (2001) went further down this path. Nagel (1995) used a simple iterated-types model to analyze her "beauty contest data," later done a little more formally by Ho, Camerer, and Weigelt (1998) and Nagel et al. (1999).

The roots of the thinking steps approaches can be found in Harsanyi's "tracing procedure" (e.g., Harsanyi and Selten, 1988). Searching for an algorithm that would select a unique equilibrium (and pick a "reasonable" one), Harsanyi suggested a procedure in which players start with a prior, imagine a best response, update the prior, and so forth to convergence. The thinking steps models use the same logic but assume that different players stop cold after different numbers of steps. Thus, the thinking steps and the Harsanyi model are related much as modern learning theories of fictitious play are related to the original incarnation (e.g., Chapter 6).

5.7.2 Payoff-Sensitive Noisy Iteration

The multiple-types idea does not relate the *degree* to which dominance is violated to the payoffs from strategies, which undoubtedly matters. For example, there is much more evidence against weak dominance than against strict dominance, probably because playing a weakly dominated strategy is a smaller mistake (in expected payoff terms). One way to allow limited reasoning that is payoff sensitive is to assume players form iterated beliefs and optimize (with noise), but add more and more noise at each step of iteration.

Capra (1999) was the first to suggest a way to do this and test it. Her idea was later modified and cleverly refined by Goeree and Holt (1999). They showed examples in which "increasing doubt" converges to quantal response equilibria (QRE). QRE has the right basic ingredients for explaining limited iterated dominance. Here's why: When λ, the responsive sensitivity parameter in QRE, is 0 then people choose randomly (they are "level-0 types") and will sometimes violate dominance. As λ rises, people are less and less likely to violate dominance, because dominated strategies will have low expected payoffs for any beliefs. The equilibrium assumption effectively builds in iterated deletion of dominated strategies as λ rises, because players come to believe that others are unlikely to violate dominance (if λ is large enough), that others are unlikely to play strategies that are dominated when dominated strategies are eliminated (when λ is larger still), and so forth. Then λ is a handy single-parameter index of the degree of limited iterated reasoning.

5.7.3 QRE Refinements: Differences and Asymmetry in λ

QRE is clearly a good tool for investigating limited iterated reasoning since lowering the response sensitivity λ will generally produce equilibria that look like data from limited iteration. However, two refinements of the QRE approach should also be explored.

1. In many games, there are clear spikes in the data which correspond to discrete levels of thinking. For example, in the p-beauty contests with $p = 2/3$ (see Figure 5.1b) many people pick exactly 33, optimizing against an expected average of 50; others pick exactly 22, and so on. QRE with a single λ will not produce such spikes, but allowing a *distribution* of λ values will.
2. In many games, such as the simple Beard and Beil games, almost all players behave rationally (e.g., obeying strict dominance) but players are not always willing to bet that *others* behave rationally. (This may be a special kind of overconfidence, in which everybody thinks others are dumber than they are.) This can be explained by extending QRE to make it asymmetric, so that players are all responsive to payoffs, but players also mistakenly believe others are less responsive than they really are (see Weiszäcker, 2000). That is, a player's behavioral probabilities are given by $P(s_i)$ (based on a sensitivity λ_i) but the same player believes another player's probabilities are $\tilde{P}(s_{-i})$ based on a lower sensitivity $\tilde{\lambda}_{-i}$ (where $\lambda_i > \tilde{\lambda}_{-i}$ to reflect more doubt about another person's behavior than about one's own). Note that this produces beliefs that are *not* in equilibrium, but in my view this is a "feature" rather than a "bug" because the idea of coupling equilibrium with quantal response always seemed rather unnatural to me (at least as a theory of play during equilibration). Formally, asymmetric QRE in logit form is defined by

$$P(s_i) = \exp\left[\lambda_i \sum_{s_{-i}} \tilde{P}(s_{-i}) u_i(s_i, s_{-i})\right] \Big/ \left(\sum_{s_k} \exp\left[\lambda_i \sum_{s_{-i}} \tilde{P}(s_{-i}) u_i(s_k, s_{-i})\right]\right), \tag{5.7.1}$$

$$\tilde{P}(s_{-i}) = \exp\left[\tilde{\lambda}_{-i} \sum_{s_i} P(s_i) u_{-i}(s_i, s_{-i})\right] \Big/ \left(\sum_{s_{-k}} \exp\left[\tilde{\lambda}_{-i} \sum_{s_i} P(s_i) u_{-i}(s_i, s_{-k})\right]\right). \tag{5.7.2}$$

Weiszäcker (2000) shows that an asymmetric QRE model organizes data from many one-shot games much better than a (symmetric) QRE can. However, the empirical asymmetry between own and perceived responsiveness (λ) largely disappears in games without dominant or dominated strategies. It appears that there is something special about players' distrust that others can do very simple kinds of logical reasoning, which suggests that a more nuanced and cognitive view will ultimately prove helpful.

5.7.4 A Poisson Cognitive Hierarchy

Teck Ho, Kuan Chong, and I (Camerer, Ho, and Chong, 2001) have been working on a cognitive hierarchy (CH) theory, which is a special kind of thinking steps model but incorporates many different elements of the approaches above. We have applied it to first-period data from dozens of games and it always predicts as least as well as Nash equilibrium, and often better (see also Gneezy, 2002). The frequency of players using K steps of thinking is given by a Poisson distribution $f(k \mid \tau)$, where τ is the mean (and the variance, in Poisson distributions) of the number of thinking steps. Players using zero steps of reasoning just choose randomly or use some other heuristic. (Random choice is useful for theoretical purposes because it means all strategies are chosen with some probability, which helps link the CH approach to equilibrium refinements.) Players using $K > 0$ steps anticipate the decisions of lower-step thinkers and best-respond to the mixture of their decisions using normalized frequencies. Formally, the expected payoff of a K-step player using strategy i is given by

$$E_i(K) = \sum_{h=1}^{m_{-i}} \pi(s_i, s_{-i}^h) \cdot \left\{ \sum_{c=0}^{K-1} \left[\frac{f(c \mid \tau)}{\sum_{c=0}^{K-1} f(c \mid \tau)} \cdot P(h \mid c) \right] \right\}, \quad (5.7.3)$$

where $P(h \mid c)$ is the chance that a c-step player chooses strategy h.

The Poisson CH has some advantages. It is very easy to compute (just iterate, starting with zero-step thinkers, on a spreadsheet), easier than QRE and even easier than Nash equilibrium in some games. It is parsimonious (only one parameter, and for empirical purposes a value of τ around 1.5 works well in many games). In most games, players who do different steps of thinking will make different decisions, but the aggregate frequency of choices will be mixed by the Poisson frequencies. Therefore, even though players are all best-responding (except zero-step thinkers), the resulting mixture will look like stochastic choice. In games with mixed equilibria, for example, the resulting predicted mixtures look a lot like data (see Camerer, Ho, and Chong, 2001, for details), but players are "purifying" since most are choosing pure strategies.

The model appears to be widely applicable to games that at first blush appear to require different sorts of models. It is obviously capable of explaining limits on iterated thinking in dominance-solvable games; but optimal choices tend to exhibit cycles in mixed games, which creates predicted mixtures close to the data. It can also explain the "magic" of coordination in entry games with market capacity (Chapter 7) because higher-step types essentially "fill in gaps" in the entry function. (To guarantee more entry as capacity grows larger requires $1 + 2\tau < e^{\tau}$, or $\tau < 1.25$, which is a reasonable number.) The Poisson distribution is also relatively easy to work with theoretically. For example, when τ is very large then the normalized frequencies of lower-level types, from level k's perspective, put all the weight on type $k - 1$. This means that, as τ grows large, the model will converge to Nash equilibrium in dominance-solvable games. There may also be close theoretical relations to risk dominance and perhaps to signaling game refinements (we conjecture that CH with large τ picks out sender–receiver equilibria that are "divine" or "universally divine"). Finally, the CH model predicts interesting effects of group size in beauty contest and centipede games (Chapter 5), and in stag hunt games (Chapter 7) which appear to correspond to experimental regularities.

A key difference between the CH approach (and other types-based approaches) and the QRE and noisy introspection approaches is that the latter generate a single, typically smooth, distribution of probabilities. CH will produce spikes. We are also optimistic that the steps of thinking which can be inferred from choices will be correlated with psychological measures like response time (more steps of thinking take longer, adjusting for working memory and calculating speed; cf. Johnson and Payne, 1985).

5.8 Conclusion

Eliminating dominated strategies is an irresistible decision principle because it (weakly) guarantees a better outcome regardless of what other players do. But believing that others obey dominance is a different matter because it is a guess about both the payoffs and the rationality of others. Believing that others believe *you* obey dominance is a different matter still. As a result, the number of levels of elimination of dominated strategies that people actually perform is an interesting question—and a fundamentally empirical one. This question has essentially been answered by the studies described in this chapter. At the risk of oversimplifying, the answer appears to be two to three steps of iterated dominance (where eliminating one's own dominated strategies is counted as one step of iteration). Results in this range have been derived in simple games with only a couple of strategies and in patent race games which are partially dominance solvable. Similar results appear in games where increasing elimination of dominated strategies *reduces* payoffs to players (centipede, price competition, travelers' dilemma, "email," and fine mechanism games) and in one game where increased iterated reasoning *increases* payoffs (the dirty faces game). The sharpest evidence on iterated dominance comes from p-beauty contest games, in which players try to choose numbers close to p times the average number. This simple game has been run on dozens of subject pools, including successful professionals and readers of business publications, with quite similar results corresponding to two to three steps of iterated reasoning.

It would be wonderful if the number of steps of iterated reasoning people tend to use were a universal constant, or even systematic across people and games, but that is unlikely. However, for doing applied game theory, simply knowing that iterated dominance is limited to a couple of steps is enough to improve predictions substantially. There is also no shame in making a prediction that is a range or subset of possible outcomes—set-theoretic predictions are common in cooperative game theory—since it is usually better to be approximately right than precisely wrong.

There are several ways limited iterated reasoning might be modeled formally.

Finally, it is important to note that equilibration does occur over time in the games described in this chapter, bringing players closer to playing dominance-solvable equilibrium. The "email game" is the most striking example of this: In the space of an hour or two, subjects can learn to play a Nash equilibrium that they themselves regarded as utterly bizarre in the early stages of the experiment. Thus, for predicting the time path of play in an experiment, or a naturally occurring process of trial-and-error learning, any theory of limited iterated reasoning will have to be coupled with a theory of learning (see Chapter 6).

Appendix: Raw Choices in Email Game and Additional Data

Tables 5.23–5.27 help you see the association between losses and strategy changes (learning). A capital letter denotes a strategy component (i.e., a message-specific choice of A or B) that was switched from the previous sequence. Underlining denotes a strategy component that earned a negative total payoff. If subjects switch only after losses, all capital letters will be underlined. If losses always create switches, the only letters that are underlined will be capitalized.

Table 5.23. *Sample feedback sheet, email game (player 1 strategy abbbbbb)*

Matrix	Last message sent by: You	Opponent	Your decision	Opponent's decision	Number of times	× Payoff	= Subtotal
M1	0	0	A	A	30	2	60
M1	0	0	A	B	0	0	0
M2	1	0	B	A	28	−4	−112
M2	1	0	B	B	0	2	0
M2	1	2	B	A	8	−4	−32
M2	1	2	B	B	3	2	6
M2	3	2	B	A	7	−4	−28
M2	3	2	B	B	5	2	10
M2	3	4	B	A	0	−4	0
M2	3	4	B	B	5	2	10
M2	5	4	B	A	1	−4	−4
M2	5	4	B	B	4	2	8
M2	5	6	B	A	0	−4	0
M2	5	6	B	B	2	2	4
M2	7	6	B	A	1	−4	−4
M2	7	6	B	B	1	2	2
M2	7	8	B	A	0	−4	0
M2	7	8	B	B	2	2	4
M2	9	8	B	A	0	−4	0
M2	9	8	B	B	1	2	2
M2	9	10	B	A	0	−4	0
M2	9	10	B	B	0	2	0
M2	11	10	B	A	1	−4	−4
M2	11	10	B	B	1	2	2
M2	11	12	B	A	0	−4	0
M2	11	12	B	B	0	2	0
						Total payoff =	−76

Source: Camerer (unpublished).

Table 5.24. *Raw choices in email game: Session 1*

	Choices for last message sent			Choices for last message sent	
	Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10		Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10
Set	Subject 1	Subject 4	Set	Subject 3	Subject 6
1	abbbbbb	abbbbb	1	bbbbbbb	abbbbb
2	abbbbbb	abbbbb	2	AAbbbbb	Bbbbbb
3	abbbbbb	abbbbb	3	aabbbbb	AAbbbb
4	aAbbbbb	abbbbb	4	aabbbbb	aBbbbb
5	aBbbbbb	aAbbbb	5	aabbbbb	abbbbb
6	abAbbbb	aabbbb	6	aabbbbb	abbbbb
7	aABbbbb	aabbbb	7	aabbbbb	abbbbb
8	aBAabbb	aabbbb	8	aabbbbb	aAbbbb
9	abBBbbb	aaAbbb	9	aabbbbb	aabbbb
10	aAAbbbb	aaBbbb	10	aaAbbbb	aabbbb
11	aBBbbbb	aaAbbb	11	aaabbbb	aabbbb
12	aAAbbbb	aaabbb	12	aaabbbb	aabbbb
13	aBBbbbb	aaabbb	13	aaaAbbb	aabbbb

Set	Subject 2	Subject 5
1	abbbbbb	ababab
2	aAbbbbb	abBbBb
3	aabbbbb	aabbbb
4	aabbbbb	aAbbbb
5	aabbbbb	aabbbb
6	aabbbbb	aabbbb
7	aaAbbbb	aaAbbb
8	aaabbbb	aaBbbb
9	aaabbbb	aBbAbb
10	aaabbbb	aAbBbb
11	aaabbbb	aaAbbb
12	aaabbbb	aaabbb
13	aaabbbb	aaabbb

Source: Camerer (unpublished).

Table 5.25. *Raw choices in email game: Session 2*

	Choices for last message sent			Choices for last message sent	
	Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10		Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10
Set	Subject 1	Subject 5	Set	Subject 3	Subject 7
1	aabbbbb	aabbbb	1	abbbbbb	abbbbb
2	aabbbbb	aabbbb	2	abbbbbb	abbbbb
3	aabbbbb	aBbbbb	3	abbbbbb	abbbbb
4	aabbbbb	aAbbbb	4	aAbbbbb	abbbbb
5	aabbbbb	aabbbb	5	aabbbbb	aAbbbb
6	aaAbbbb	aabbbb	6	aaAbbbb	aBbbbb
7	aaabbbb	aaAbbb	7	aaBbbbb	aAbbbb
8	aaabbbb	aaBbbb	8	aaAbbbb	aabbbb
9	aaabbbb	aaAbbb	9	aaabbbb	aaAbbb
10	aaaAbbb	aaabbb	10	aaabbbb	aaabbb
11	aaaabbb	aaabbb	11	aaabbbb	aaabbb
12	aaaabbb	aaabbb	12	aaaAbbb	aaabbb
13	aaaabbb	aaabbb	13	aaaabbb	aaaAbb
14	aaaaAbb	aaaAbb	14	aaaabbb	aaaabb
15	aaaaabb	aaaBbb	15	aaaabbb	aaaabb
Set	Subject 2	Subject 6	Set	Subject 4	Subject 8
1	abbbbbb	abbbbb	1	bbbbbbb	abbbbb
2	aAbbbbb	abbbbb	2	AAbbbbb	abbbbb
3	aabbbbb	abbbbb	3	aabbbbb	aAbbbb
4	aabbbbb	aAbbbb	4	aabbbbb	aBbbba
5	aabbbbb	aabbbb	5	aabbbbb	aAbbbb
6	aabbbbb	aabbbb	6	aaAbbbb	aBbbbb
7	aabbbbb	aBAbbb	7	aaBbbbb	aAbbbb
8	aabbbbb	aABbbb	8	aaAbbbb	aBbbbb
9	aabbbbb	aabbbb	9	aaabbbb	aAAbbb
10	aabbbbb	aabbbb	10	aaabbbb	aaabbb
11	aaAbbbb	aabbbb	11	aaabbbb	aaabbb
12	aaabbbb	aaAbbb	12	aaabbbb	aaBbbb
13	aaabbbb	aaabbb	13	aaabbbb	aaAbbb
14	aaaAbbb	aaabbb	14	aaaAbbb	aaabbb
15	aaaabbb	aaaAbb	15	aaaabbb	aaabbb

Source: Camerer (unpublished).

Table 5.26. *Raw choices in email game: Session 3*

	Choices for last message sent			Choices for last message sent	
	Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10		Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10
Set	Subject 1	Subject 5	Set	Subject 3	Subject 7
1	aabbbbb	abbbbb	1	aabbbbb	abbbbb
2	aabbbbb	abbbbb	2	aabbbbb	aAbbbb
3	aabbbbb	abbbbb	3	aabbbbb	aabbbb
4	aabbbbb	aAbbbb	4	aabbbbb	aabbbb
5	aabbbbb	aabbbb	5	aaAbbbb	aabbbb
6	aaAbbbb	aabbbb	6	aaabbbb	aabbbb
7	aaabbbb	aaAbbb	7	aaabbbb	aaAbbb
8	aaaAbbb	aaabbb	8	aaabbbb	aaabbb
9	aaaabbb	aaaAbb	9	aaabbbb	aaabbb
10	aaaaAbb	aaaabb	10	aaaAbbb	aaaAbb
11	aaaaabb	aaaaAb	11	aaaaAbb	aaaaAb
12	aaaaaAb	aaaaaA	12	aaaaaAb	aaaaab
13	aaaaaaA	aaaaaa	13	aaaaaaA	aaaaaA
14	aaaaaaa	aaaaaa	14	aaaaaaa	aaaaaa
Set	Subject 2	Subject 6	Set	Subject 4	Subject 8
1	aabbbbb	abbbbb	1	abbbbbb	abbbbb
2	aabbbbb	abbbbb	2	aAbbbbb	aAbbbb
3	aabbbbb	abbbbb	3	aabbbbb	aBbbbb
4	aabbbbb	abbbbb	4	aabbbbb	aAbbbb
5	aaAbbbb	abbbbb	5	aabbbbb	aabbbb
6	aaabbbb	abbbbb	6	aaAbbbb	aabbbb
7	aaabbbb	aAbbbb	7	aaabbbb	aaAbbb
8	aaabbbb	aabbbb	8	aaabbbb	aaabbb
9	aaaAbbb	aaAbbb	9	aaaAbbb	aaabbb
10	aaaaAbb	aaaAbb	10	aaaaAbb	aaaAbb
11	aaaaabb	aaaaAb	11	aaaabb	aaaaAb
12	aaaaaAb	aaaaaA	12	aaaaaAb	aaaaab
13	aaaaaaA	aaaaaa	13	aaaaaaA	aaaaaA
14	aaaaaaa	aaaaaa	14	aaaaaaa	aaaaaa

Source: Camerer (unpublished).

Table 5.27. Raw choices in email game: Session 4

	Choices for last message sent			Choices for last message sent	
	Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10		Player 1 messages 0,1,3,5,7,9,11	Player 2 messages 0,2,4,6,8,10
Set	Subject 1	Subject 5	Set	Subject 3	Subject 7
1	aabbbbb	abbbbb	1	aabbbbb	abbbbb
2	aabbbbb	abbbbb	2	aabbbbb	abbbbb
3	aabbbbb	aAbbbb	3	aabbbbb	aAbbbb
4	aabbbbb	aabbbb	4	aaAbbbb	aabbbb
5	aaAbbbb	aabbbb	5	aaabbbb	aabbbb
6	aaabbbb	aabbbb	6	aaabbbb	aabbbb
7	aaabbbb	aaAbbb	7	aaaAbbb	aaAbbb
8	aaaAbbb	aaabbb	8	aaaabbb	aaabbb
9	aaaaAbb	aaaAbb	9	aaaabbb	aaaAbb
10	aaaaabb	aaaabb	10	aaaaAbb	aaaaAb
11	aaaaaAb	aaaaAb	11	aaaaabb	aaaaab
12	aaaaaaA	aaaaab	12	aaaaaAb	aaaaaA
13	aaaaaaa	aaaaaA	13	aaaaaaA	aaaaaa
14	aaaaaaa	aaaaaaa	14	aaaaaaa	aaaaaa
Set	Subject 2	Subject 6	Set	Subject 4	Subject 8
1	aabbbbb	abbbbb	1	aabbbbb	abbbbb
2	aAbbbbb	abbbbb	2	aabbbbb	abbbbb
3	aabbbbb	aAbbbb	3	aabbbbb	aAbbbb
4	aaAbbbb	aabbbb	4	aaAbbbb	aabbbb
5	aaabbbb	aabbbb	5	aaaAbbb	aabbbb
6	aaabbbb	aaAbbb	6	aaaBbbb	aabbbb
7	aaabbbb	aaabbb	7	aaaAbbb	aaAbbb
8	aaaAbbb	aaabbb	8	aaaabbb	aaabbb
9	aaaabbb	aaaAbb	9	aaaaAbb	aaaAbb
10	aaaaAbb	aaaaAb	10	aaaaabb	aaaabb
11	aaaaaAb	aaaaab	11	aaaaaAb	aaaaAb
12	aaaaaaA	aaaaab	12	aaaaaab	aaaaab
13	aaaaaaa	aaaaaA	13	aaaaaaA	aaaaaA
14	aaaaaaa	aaaaaa	14	aaaaaaa	aaaaaa

Source: Camerer (unpublished).

6
Learning

THE QUESTION OF HOW AN EQUILIBRIUM ARISES in a game has been largely avoided in the history of game theory, until recently. Equilibrium concepts implicitly assume that players either figure out what equilibrium to play by reasoning, follow the recommendation of a fictional outside arbiter (if that recommendation is self-enforcing), or learn or evolve toward the equilibrium.

Most of the research on learning and evolution is theoretical, but it is unlikely that theorizing alone will explain how people learn without the input of empirical regularity and careful testing. Theories generally use the mathematics of stochastic processes to prove theorems about the limit properties of different rules (see Weibull, 1995, and Fudenberg and Levine, 1998, for overviews). Rules with plausible limiting behavior are thought to be better descriptions. However, if limiting behavior takes months, years, or decades to unfold, then limit theorems are not as useful as being able to predict the path of equilibration.

This chapter focuses on using experimental data to test models of learning. Note that this chapter is quite different from the others in this book. The focus is not regularity in a class of games; the focus is on what types of model generally fit the path of learning, and why. Learning is defined as an observed change in behavior owing to experience. Statistical models of learning therefore predict how probabilities of future choices are affected by historical information.

Experimental data are a good way to test models of learning because control over payoffs and information means we can be sure what subjects know (and know others know, and so on), what they expect to earn from different strategies, what they have experienced in the past, and so forth.

Table 6.1. *Stag hunt game*

Row player strategies	L	R
T	3,3	0,1
B	**1,0**	1,1

Since most models make detailed predictions that require this information, laboratory control is indispensable for sorting out candidate models at this early stage in the research. Laboratory pre-screening can identify models that are likely to work well in naturally occurring situations such as auctions, firm choices of prices and quantities, bargaining outcomes in strikes and divorce, and consumer learning about preferences.[1]

6.1 Theories of Learning

There are many approaches to learning in games: evolutionary dynamics; reinforcement learning; belief learning; sophisticated (anticipatory) learning; experience-weighted attraction (EWA) learning; imitation; direction learning; and rule learning.

The stag hunt game in Table 6.1 (see also Chapter 7) helps show exactly how some of the theories work. Let's focus only on learning by the row player. Suppose in period t the row player chose B and the column player chose L, yielding payoffs of 1 and 0 respectively for row and column (printed in bold in the table). The forgone payoff from row's unchosen strategy (T) is 3.

Many theories assume strategies have numerical evaluations, which we call "attractions." Learning rules can be characterized by how attractions are updated in response to experience. Denote the attractions for strategies T and B *before* the period t play by $A^{\mathrm{T}}(t-1)$ and $A^{\mathrm{B}}(t-1)$. Attractions are mapped into predicted choice probabilities using some statistical rule (usually logit or power). Table 6.2 summarizes how attractions for T and B are predicted to change according to various theories.

1. *Evolutionary approaches* assume a player is born with a strategy and plays it, usually in random matching with members of a population.

[1] A great field application is Weisbuch, Kirman, and Herreiner (2000), who apply a reinforcement model to the development of trading relationships in the Marseilles fish market. Ho and Chong (in press) predict consumer learning about products such as ice cream and analgesics. (Their model predicts better out of sample by a substantial margin than the leading model in marketing research, and has 80 percent *fewer* parameters.)

Table 6.2. *How learning theories update attractions in the stag hunt example*

		Attraction for . . .	
Theory	Functional form or statistic	B in $t+1$ $A^{\mathrm{B}}(t)$	T in $t+1$ $A^{\mathrm{T}}(t)$
Reinforcement	Averaging	$\phi A^{\mathrm{B}}(t-1)+(1-\phi)(1-\epsilon)(1-\rho(t-1))$	$\phi A^{\mathrm{T}}(t-1)+(1-\phi)\epsilon(1-\rho(t-1))$
	Cumulative	$\phi A^{\mathrm{B}}(t-1)+(1-\epsilon))(1-\rho(t-1))$	$\phi A^{\mathrm{T}}(t-1)+\epsilon(1-\rho(t-1))$
Belief learning	Beliefs	$P_t(L)=\dfrac{3\rho+1}{5\rho+1}$	$P_t(R)=\dfrac{2\rho+0}{5\rho+1}$
	Attractions	1	$\dfrac{3(3\rho+1)+0(2\rho+0)}{5\rho+1}$
EWA		$\dfrac{\phi N(t-1)A^{\mathrm{B}}(t-1)+1}{\rho N(t-1)+1}$	$\dfrac{\phi N(t-1)A^{\mathrm{T}}(t-1)+3\delta}{\rho N(t-1)+1}$
Replicator dynamics	Population proportions	$\dfrac{p^{\mathrm{T}}(t+1)}{p^{\mathrm{T}}(t)}=1+\alpha(3p^{\mathrm{L}}(t)-1)(1-p^{\mathrm{T}}(t))$	$\dfrac{p^{\mathrm{B}}(t+1)}{p^{B}(t)}=1+\alpha(1-3p^{\mathrm{L}}(t))(1-p^{\mathrm{B}}(t))$

Strategies that are relatively successful increase the player's fitness—perhaps they provide food or avoid attacks by predators—and enable players to survive longer or reproduce more frequently. Mathematical exploration of evolutionary models has been a hot topic of research recently (see, e.g., Weibull, 1995). I will discuss them in this chapter only briefly because evolutionary models apply best to animals with genetically heritable strategies, or to human cultural evolution (Boyd and Richerson, 1985), neither of which explains rapid individual learning in the lab.

2. One step up from evolutionary models in the cognitive sophistication that agents are assumed to have are *reinforcement* approaches (also called stimulus-response or rote learning). Choice reinforcement assumes strategies are "reinforced" by their previous payoffs. Reinforcement may also "spill over" to strategies that are similar to the chosen strategy (e.g., neighboring strategies, if strategies are rank ordered). Reinforcement learning is a reasonable theory for players with very imperfect reasoning ability (nonhuman animals pecking levers in the lab or foraging in the wild) or for human players who know absolutely nothing about the forgone or historical payoffs from strategies they did not choose.[2]

   Applied to our example, reinforcement theories update attractions according to $A^{\mathrm{B}}(t) = \phi A^{\mathrm{B}}(t-1) + (1-\epsilon)1 - \rho(t-1)$ and $A^{\mathrm{T}}(t) = \phi A^{\mathrm{T}}(t-1) + \epsilon$, if the attractions "cumulate." The parameter ϵ represents a spillover or generalization of reinforcement from one strategy (B) to neighboring strategies (T). (Which strategies are "neighboring" is of course an empirical question, having to do with the psychology of categorization.[3]) Note that, when $\epsilon = 0$, strategy T receives no reinforcement. In a variant of this model, attractions are weighted averages rather than cumulations, so $A^{\mathrm{B}}(t) = \phi A^{\mathrm{B}}(t-1) + (1-\phi)(1-\epsilon)$ and $A^{\mathrm{T}}(t) = \phi A^{\mathrm{T}}(t-1) + (1-\phi)\epsilon$.

3. *Belief learning models* assume players update beliefs about what others will do based on history, and use those beliefs to determine which

[2] There is a closer relation between evolutionary models of population learning and models of individuals learning by reinforcement than one might suspect at first glance. Börgers and Sarin (1997) show that in the Cross (1973) reinforcement learning model, if strategies are reinforced by an increasingly small fraction of their payoffs (to mimick a reduction in the time between iterations becoming smaller), as that fraction goes to zero the reinforcement dynamics converges to replicator dynamics.

[3] In his discussion of confirmation bias in gambling Wagenaar (1984, p. 109) gives an interesting example of how spillover may work in perhaps surprising ways: "[Consider] the roulette player who suddenly places a large single bet on number 24, completely out of his routine betting pattern. His reason was that 12 is always followed by a 24. After he lost his bet I enquired what had gone wrong. He said, 'It almost worked.' The number that did come out was 15, which is adjacent to 24 on the number wheel. Probably, he would have considered other outcomes like 5, 10, and 33 also confirmations, because they are nearby on the wheel. Also he could have taken the outcomes 22, 23, 25, and 26 as confirmations because their numerical value is close [to 24]."

strategies are best. A popular model is "fictitious play." In fictitious play, players keep track of the relative frequency with which another player has played each strategy in the past. These relative frequencies are beliefs about what that player will do in the upcoming period. Players then calculate expected payoffs for each strategy given these beliefs, and choose strategies with higher expected payoffs more frequently.

Fictitious play counts all previous observations equally. At the opposite extreme is Cournot best-response dynamics: Assume the strategy played most recently by others will be played again. Cournot dynamics weight the most recent past very heavily and dismiss or discard older experiences.

Weighting distant experiences less than recent ones gives a hybrid form called "weighted fictitious play" (Cheung and Friedman, 1997; Fudenberg and Levine, 1998). In our example, suppose that before period t the player had a belief $P_{t-1}(L) = 0.6$, and the strength of this belief (in units of experience) is 5 (i.e., $P_{t-1}(L) = 3/5$ and $P_{t-1}(R) = 2/5$). Then in weighted fictitious play beliefs are updated according to $P_t(L) = (3\rho + 1)/(5\rho + 1)$ and $P_t(R) = (2\rho + 0)/(5\rho + 1)$, where ρ is a decay factor. The boundary case $\rho = 0$ is Cournot best-response dynamics and $\rho = 1$ is fictitious play (keep a running count of the number of times L and R are chosen and take the ratio of L observations to total observations to be the belief $P_t(L)$). Beliefs are then used to compute expected payoffs, so $A^{\mathrm{T}}(t) = [3(3\rho + 1) + 0(2\rho + 0)]/(5\rho + 1)$. Notice that the payoff the row player actually received in period t plays no special role in determining choice behavior in period $t + 1$.

4. *Experience-weighted attraction (EWA) learning* was designed by myself and Teck Ho (1999a) to combine the most appealing elements of reinforcement and weighted fictitious play in a hybrid or "gene-splice." (The Nobel laureate Francis Crick, who discovered DNA, wrote: "In nature a hybrid species is usually sterile; in science the opposite is often true" [1988, p. 150].)

   The model adds a key feature to reinforcement and belief learning: δ, the weight players give to forgone payoffs from unchosen strategies. When parameters are restricted to have certain values, EWA reduces to a simple version of choice reinforcement in which only chosen strategies are reinforced. When parameters are restricted in a different way, EWA reduces *exactly* to weighted fictitious play. So EWA is a *family* of learning rules with reinforcement and belief learning as extreme cases.

   In our example, EWA attractions are updated according to $A^{\mathrm{B}}(t) = [\phi N(t-1)A^{\mathrm{B}}(t-1) + 1)/(\phi(1-\kappa)N(t-1) + 1]$ and $A^{\mathrm{T}}(t) = [\phi N(t-1)A^{\mathrm{T}}(t-1) + 3\delta]/[\phi(1-\kappa)N(t-1) + 1]$. Notice that, if $\delta > 1/3$, a player could reinforce T more strongly than B even though T was not chosen. Players are then predicted to shift choice probability away from B and toward T. Notice that the choice of the column player plays a role in the row player's updating (because it determines the forgone payoff), and whether a payoff was received or simply imagined also matters (since the forgone payoff is weighted by δ).

5. In adaptive models such as fictitious play and EWA, players only look back at what other players have done previously. As a result, adaptive players will never shift their beliefs to expect something that they did not expect or observe before, and adaptive players will ignore information about other players' payoffs. But experiments that vary whether players know other players' history and payoffs show that players do care about what other players earn.

   Both limitations are overcome in models with *anticipatory learning* or *sophistication* (e.g., Milgrom and Roberts, 1991; Selten, 1986; Camerer, Ho, and Chong, 2002a; Stahl, 1999a). In these models, players do use information about other players' payoffs to reason more thoughtfully about what other players will do in the future.[4] These sophisticated models are a kind of belief learning because players *are* forming beliefs and best-responding according to them; it's just that their beliefs are more sophisticated than merely guessing that players will repeat their past choices.

   In the stag hunt example, adding sophistication means the row player knows that the column player earned 0, and knows the column player could have earned 1 from choosing R. If the row player thinks the column player is learning according to Cournot, therefore, she will predict that column will choose R. A sophisticated row player will therefore choose B, expecting a payoff of 1 (from column's R choice). Note that sophistication requires knowledge of the other player's payoffs (to compute her likely response). Including sophistication is also a way to model the fact that players behave differently when they are randomly rematched and when they are matched with "partners" repeatedly (see Camerer, Ho, and Chong, 2002a,b, and Chapter 8).
6. Players sometimes learn by *imitating* the strategies of others. Imitation could be independent of payoffs or could depend on payoffs—for example, players imitate the most successful player they see.
7. In *learning direction theory*, a player determines her ex post best response and adjusts her previous strategy in the direction of the best response (Selten and Stöcker, 1986). Learning direction theory combines the idea of moving toward the best response from Cournot-style

[4] Another example is an adaptive model in which players are assumed to detect which strategies are dominated for other players, and think others will never play those strategies (Cooper, Garvin, and Kagel, 1997b).

belief learning and the idea that one anchors on the previous strategy from reinforcement learning or habit models. It has not been generally defined for games without ordered strategies.

8. *Rule learning* assumes that people have decision rules that map histories into strategy choices. They learn about which *rules* to use, rather than about which specific *strategy* to use. Possible rules could include those listed above, and other rules such as tit-for-tat, level-k reasoning (see Chapter 5), and so forth.

   To illustrate using stag hunt, suppose the rules a player is considering are minimax (choose the strategy with the highest minimum payoff, which prescribes B) and a level-1 rule (best-respond to a random choice by others), which prescribes T. Given that the column player chose L, minimax would be reinforced by the payoff to the strategy it recommended (B), which is 1, and the level-1 rule would be reinforced by 0.

Most of the theories above are plausible, so it is useful to have a clear set of criteria for judging which theories are best (and for which purposes).

One way to judge plausibility is to ask what sorts of information are used by the updating rules. If there is information a theory does not require, which people use when it is available, the theory is incomplete. Theories that demand more information than people are able to integrate effectively are dubious too.

Notation is necessary to proceed further. Denote player i's jth strategy (out of m_i strategies) by s_i^j. The strategies that i and all other players (denoted $-i$) actually chose in period t are denoted $s_i(t)$ and $s_{-i}(t)$, respectively. Player i's payoff to playing s_i^j when others played s_{-i}^k is $\pi_i(s_i^j, s_{-i}^k)$. Player i's (ex post) best response is $b_i(s_{-i}(t)) \equiv \text{argmax}_k \ \pi_i(s_i^k, s_{-i}(t))$.

Table 6.3 shows the minimal information players need in different theories.[5] (Evolutionary theories are not listed because, taken seriously, they assume players use no information at all since they are born with fixed strategies.)

Actual information use depends on what information is available and on players' cognitive capacities. In high-information conditions, such as most of the experiments discussed in this chapter, all the information listed in Table 6.3 is available to players. Theories that assume players do not use

[5] Some kinds of information are sometimes necessary to compute other information. For example, if the payoffs are known then the choice by other players $s_{-i}(t)$ is needed to compute player i's received payoff $\pi_i(s_i(t), s_{-i}(t))$. However, it is possible to know player i's received payoff without knowing what other players did. Since Table 6.3 shows the *minimal* information needed by theories, a theory that requires knowing only the received payoffs (such as reinforcement) would *not* require knowledge of the choice $s_{-i}(t)$.

Table 6.3. *Minimal information used by various learning theories*

Information	Learning theories						
						Imitate the . . .	
	Reinforcement	Beliefs	Direction learning	EWA	Sophistication	Average	Best
i's choice $s_i(t)$	X		X	X			
$-i$'s choice $s_{-i}(t)$		X		X	X	X	X
i's received payoff $\pi_i(s_i(t), s_{-i}(t))$	X	X		X			
i's forgone payoffs $\pi_i(s_i^j, s_{-i}(t))$		X		X	X		
i's best response $b_i(s_{-i}(t))$			X				
$-i$'s received payoff $\pi_{-i}(s_i(t), s_{-i}(t))$					X		X
$-i$'s forgone payoffs $\pi_i(s_i^j, s_{-i}^k)$					X		

all that information are implicitly assuming some constraint on cognitive capacity.

Different theories can be tested indirectly, by seeing whether behavior changes when required information is or is not available, or by measuring what information players look at. The few studies of this type show that subjects learn faster when they have full payoff information (which contradicts reinforcement), look up information about their own previous payoffs (which contradicts belief learning), and behave differently when they know the payoffs of others (implying sophistication).[6]

Another way to judge learning rules is to ask which rules would survive in an evolutionary competition (e.g., Heller, 1998). (The key to modeling this competition effectively is to specify the costs of rules of different complexity.) Josephson (2001) found that, in two-player games, EWA rules with high values of δ ("vivid imagination") tended to persist.

It is useful to list other properties we generally like theories to have. Here is one such list of desiderata: fit, coherence (or surprisingness), fruitfulness, and analytical tractability. Good theories *fit* and *predict* data well (adjusting for degrees of freedom, of course, to guard against overfitting). Good theories make *coherent* sense of unrelated phenomena; when their relation is *surprising* that's a bonus.[7] A theory that can be easily understood and applied by a broad body of scholars, and that provides a large body of solvable puzzles, is *fruitful*. And a theory should be *"analytically tractable"*—clear and parsimonious enough that theorists can prove something about what follows from assuming the theory is true. As you read, judge for yourself which theories are strongest on these criteria (or any criteria you like).

The empirical literature on fitting learning models to experimental data has grown dramatically in just a few years. I will first discuss studies that test only one approach, then discuss comparative studies.

6.2 Reinforcement Learning

Reinforcement approaches are derived from behaviorist psychology. Behaviorism was an extreme and important chapter in the history of psychology from about 1920 to 1960. The behaviorists were fed up with vague "mentalist" accounts of thinking processes that could not be directly observed. So they imposed an intellectual prohibition on speculation about cognitive

[6] Mookerjhee and Sopher (1994), Van Huyck, Battalio, and Rankin (2001a), and Rapoport and Erev (1998) find different learning when forgone payoff information is available. Partow and Schotter (1996) find differences when other players' payoffs are known.

[7] A theory that affirms the obvious adds nothing new to knowledge; in Bayesian terms, we seek theories that create posterior beliefs that are far from priors.

processes and insisted that all behavior could be explained as learned responses to previous reinforcement. Although the behaviorists left behind an invaluable legacy of careful experimentation (which had not existed in psychology before), their basic framework has been largely abandoned since then because it could not explain how people learned vicariously, and so quickly, in domains with a vast array of possible responses and no direct reinforcement (e.g., children learning language). And cognitive constructs were increasingly added on to explain these phenomena,[8] reinventing the mentalist ideas that behaviorists eschewed in the first place. Note that behaviorism was never "disproved." It was largely *replaced* by fruitful analogies between the brain and the computer, and the brain and "connectionist" (parallel distributed processing, or PDP) neural networks.[9]

There are three waves of formal research on reinforcement learning in decision and game theory. Fifty years ago Bush and Mosteller (1955) and others formalized simple reinforcement rules and applied them to learning in decisions. In a one-person second wave, Cross (1973, 1983) applied reinforcement to economic decisions. Unfortunately, his pioneering work was largely ignored until a third wave, about ten years later, when Arthur (1991, 1994) revived the reinforcement approach and applied it to simple decisions. McAllister (1991), Mookerjhee and Sopher (1994, 1997), Roth and Erev (1995), Sarin and Vahid (2001), and others later applied reinforcement to games.

6.2.1 Reinforcement in Weak-Link Games

In McAllister's (1991) approach, rewards are normalized within payoff bounds, and updated according to

$$A_i^j(t) = \phi A_i^j(t-1) + (1-\phi)\pi_i(s_i(t), s_{-i}(t))_j \qquad s_i^j = s_i(t)_j \qquad (6.2.1)$$

$$A_i^j(t) = \phi A_i^j(t-1)_j \qquad \text{otherwise.} \qquad (6.2.2)$$

Using an indicator function $I(x, y)$ which equals 1 when $x = y$ and 0 otherwise, the two rules can be written as

$$A_i^j(t) = \phi A_i^j(t-1) + (1-\phi) I(s_i^j, s_i(t))\pi_i(s_i(t), s_{-i}(t)). \qquad (6.2.3)$$

[8] An example is "stimulus generalization," the indirect reinforcement of one stimulus by a different stimulus which is similar.

[9] Reinforcement is still widely used to study animal learning and some domains of human behavior, such as treatment of phobias, which are widely thought to tap "old brain" mechanisms for which simple animal learning rules may also apply to humans. It is also useful for explaining learning in simple decision environments such as slot machine gambling (e.g., Lea, Tarpy, and Webley, 1987, p. 287), where partial reinforcement is clearly a powerful way to keep people pulling the lever.

Table 6.4. *Simulated reinforcement learning of weak-link data*

Action	Period 1 data	Period 8 data	Period 8 simulated
1	0.019	0.654	0.25
2	0.047	0.252	0.50
3	0.047	0.028	0.13
4	0.168	0.028	0.06
5	0.318	0.009	0.06
6	0.093	0.000	0.00
7	0.308	0.028	0.00

Source: McAllister (1991).

The probability $P_i^j(t)$ of choosing strategy s_i^j in period t is updated according to

$$P_i^j(t) = (P_i^j(t) - \alpha \cdot \rho_j(t) P_i^j(t)) \Big/ \left(\sum_{k=1}^{m_i} P_i^k(t) - P_i^k(t)\alpha \cdot \rho_k(t) \right), \quad (6.2.4)$$

where α is an adjustment factor which affects the rate of learning and $\rho_j(t)$ is an adjustment ratio. Learning is faster when α is high and γ is low.

McAllister applies this model to the weak-link coordination game data of Van Huyck, Battalio, and Beil (1990) (see Chapter 7). It learns too slowly unless attractions are updated according to either actual or forgone payoffs (i.e., $A_i^j(t) = \phi A_i^j(t-1) + (1-\phi)\pi_i(s_i^j, s_{-i}(t))$). (Like others who came later, McAllister did not realize that this important change makes the reinforcement model a close relative of belief learning.) Experimenting with several parameter values, he finds that for fast updating—$\alpha = 0.75$, $\gamma = 0.5$, and $\phi = 0.5$—simulated paths converge toward the observed data in eight periods (see Table 6.4), although actual players converge more strongly to strategy 1 than the simulations do.

6.2.2 *Reinforcement with Payoff Variability*

Roth and Erev (1995) ask whether reinforcement learning can explain differences among three games—ultimatums, ultimatums with proposer competition ("market games"), and best-shot public goods games. In all three games the division of surplus is predicted to be extremely uneven. In experiments, however, subjects converge more strongly to uneven divisions in market games and best-shot games, and converge to nearly-equal offers in ultimatum games (see Chapter 2). Roth and Erev wanted to see whether learning models could reproduce these patterns (which are also parsimoniously explained by new models of social preference, see Chapter 2).

In the simplest form of their model, attractions cumulate according to

$$A_i^j(t) = \phi \cdot A_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t)), \tag{6.2.5}$$

and are mapped into probabilities using a power function (with $\lambda = 1$). The same model was proposed earlier by Harley (1981), who suggested a neural interpretation in which reinforcements are levels of brain activation.

Cumulating attractions with a power probability function means later payoffs have less proportional impact than early ones, so learning slows down over time.[10] In earlier models (Bush and Mosteller, 1955; Cross, 1983), choice probabilities are updated directly by reinforcements, and learning slows down because an updating parameter is assumed to fall over time (Sarin, 1995).

Their model also has a scaling parameter $S(1)$ (which fixes the sum of the initial attractions to be $S(1)$ times the average game payoff) and a local experimentation parameter ϵ, which spills reinforcement of a chosen strategy over to its two neighboring strategies.[11]

Like McAllister, Roth and Erev fit the model by choosing small sets of parameter values, simulating model behavior for those values, and comparing the simulations with data informally. (Later studies choose finer-grained parameter values using a fit-maximizing technique that permits statistical tests.)

When initial conditions are matched to the data, the simulated learning paths in ultimatum games move in the right direction, but much more slowly than subjects do. In the best-shot games, players 1 and 2 choose a public good contribution 0, 2, or 4 (denoted q_1 and q_2), which is privately costly. The public good is an increasing function of the *maximum* of q_1 and q_2 and benefits both players equally. Given the payoffs, the subgame perfect prediction is $q_1 = 0$ and $q_2 = 4$.

Table 6.5 reports changes in the probabilities that q_1 is 0 or 2, and probabilities of q_2 being chosen as 0 or 2 in response to the perfect equilibrium choice $q_1 = 0$ (for the partial information condition). For player 1 choices of q_1, the simulated change from period 1 to 100 is close to the actual 10-

[10] The "power law of practice" refers to the fact that, in tasks such as adding numbers, recognizing previously presented sentences, or rolling cigars, the speed of performance or recall reliably obeys a power law of the amount of experience N, $P = ae^{-bN}$ (e.g., Anderson, 2000, p. 187). Of course, since the learning environment in a game is nonstationary (owing to learning by others), an exact power law usually won't hold, but the "power law" phrase is used informally to refer to a slowdown in learning.

[11] They also have a cutoff parameter μ, which sets probabilities below μ equal to 0.

Table 6.5. Data and simulated reinforcement behavior in partial information best-shot public goods games

	Data change	Simulated change	
Statistic	Periods 1–10	Periods 1–10	Periods 1–100
$\Delta P(q_1=0)$	+0.48	+0.22	+0.55
$\Delta P(q_1=2)$	−0.20	−0.10	−0.22
$\Delta P(q_2=0\mid q_1=0)$	+0.11	−0.05	−0.09
$\Delta P(q_2=2\mid q_1=0)$	−0.35	+0.04	+0.08

Source: Numbers estimated from Roth and Erev (1995), Figure 4.
Note: $\Delta P(q_1=i)$ represents change in the proportion of player 1s choosing i over the periods indicated in the column headings. $\Delta P(q_2=k\mid q_1=i)$ represents change in the proportion of player 2s choosing k in response to a q_1 choice of i.

period change, but the simulated 10-period change is only half as large as it is in the data. However, the simulations do not capture the change in player 2 responses well.

In the market games, nine Proposers offer a division of a fixed pie to a single Responder, who takes the best offer. Offers converge very rapidly, in just one or two periods, to offering almost all the points. The reinforcement model described above cannot explain the rapid rate of convergence, because eight of nine Proposers' offers are rejected, so they get no reinforcement and do not learn to offer more. To explain the rapid rate of convergence, Roth and Erev switch to a different model, in which players' unchosen strategies are reinforced by the payoff of the *winning* bidder (which is basically belief learning). They use the same winning-payoff model to speed up learning in weak-link coordination games, where simulated learning from received payoffs is too slow to match the data (see McAllister, 1991, and Roth, 1995b, pp. 37–40).

These early studies show that reinforcement learning from own payoffs can approximate the direction of learning in ultimatum and best-shot games (thought it is too slow) but does a poor job explaining learning in market and weak-link games. Given these results, an obvious direction, taken by many later authors (including myself), was to search for more robust models which do not have the empirical weaknesses of reinforcement. Another was to acknowledge that models sometimes fail badly, but to continue to explore other domains in which reinforcement may work adequately.

Pursuing the latter path, Erev and Roth (1998) applied a variety of reinforcement models to constant-sum games with mixed equilibria. They

Table 6.6. Average model fit (MSD) at individual level

	Data	Experiment		
		O'Neill (1987)	Ochs (1995a)	Erev and Roth (new data)
Basic reinforcement	1st half	0.20	0.13	0.24
	2nd half	0.18	0.12	0.21
Fictitious play	1st	0.19^{+}	0.14	0.24
	2nd	0.19^{-}	0.15^{-}	0.22
Nash equilibrium	1st	0.18^{+}	0.15^{-}	0.31^{-}
	2nd	0.18	0.14^{-}	0.30^{-}

Source: Erev and Roth (1998).
Note: Superscript − denotes significantly worse than basic reinforcement by paired t-test within-subject (one-tailed, $p = .05$); superscript + denotes significantly better.

chose games with long spans (100 or more trials) in which equilibration is slow "to observe intermediate term as well as short term behavior." They explored various specifications and levels of analysis. One comparison fit data to individual subjects using three models: basic reinforcement ($\phi = 1$, $\epsilon = 0$), a stochastic power form of fictitious play (PFP), and Nash equilibrium. The mean squared deviations (MSDs) between observed choices and predictions are shown in Table 6.6 for the first and second halves of each of three data sets. Table 6.6 shows that equilibrium predictions are not bad in absolute terms, but are significantly worse than the learning models in two of three data sets. Stochastic fictitious play (PFP) and reinforcement are about equally accurate across the three data sets (PFP is slightly less accurate for later periods).

Erev, Bereby-Meyer, and Roth (1999) add a payoff variability term to their earlier model and apply it to risky choices. The idea is to divide attractions by the variability of received payoffs as a way of slowing down learning when the environment is variable, and sharpening convergence when the environment is stable.

Roth et al. (1999) use this model to fit and predict data from randomly sampled 2×2 games with mixed-strategy equilibria. They stress that it is good to have a precise criterion for measuring how useful a theory is (see Harless and Camerer, 1994) and propose a measure called "predictive value." A theory's predictive value is the number of observations of data (new subjects or new periods) one can save by using the theory to make predictions instead of running more subjects. (It is a measure of "labor savings.")

In more recent work on their sample of 2×2 games with mixed equilibria, reinforcement, Camerer-Ho EWA, and fictitious play models are about equally accurate, and all the learning models are substantially better than equilibrium.

6.2.3 Reinforcement with "Mood Shocks"

Sarin and Vahid (2001) propose a reinforcement model in which attractions are equal to lagged attractions plus a learning rate parameter γ times the "surprise," the difference between the received payoff and the previous attraction. That is,

$$A_i^j(t) = A_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \gamma(\pi_i(s_i^j, s_{-i}(t)) - A_i^j(t-1)). \quad (6.2.6)$$

(This is just Erev and Roth's simple model with $\phi = 1$ and attractions equal to averages.) Sarin and Vahid assume that players always choose the strategy with the highest attraction, but they average across many simulations with random shocks included, so the model is observationally similar to one with stochastic choice.

Sarin and Vahid first compare frequencies of simulated choices with frequencies in block-averaged data from the games with mixed equilibria that Erev and Roth (1998) sampled. The one-parameter Sarin–Vahid model fits better than Roth–Erev's three-parameter model in seven data sets, worse in four, and slightly better overall. A subject-by-subject analysis is similar. The estimated learning parameter $\hat{\gamma}$ is 0.010, which is very small and is likely to be much larger in other classes of games. It is not clear *why* the Sarin–Vahid averaging model does modestly better than the Roth–Erev cumulation model. They attribute its relative success to the maximization assumption, but I think that's wrong because averaging simulations is noisy best response. Averaging is probably just a better description of the cognitive process in these games than cumulation is.

6.2.4 Information Conditions

Because reinforcement models assume players care only about their history of payoffs, the models can be applied in low-information and changing environments. But it also implies that players in high-information environments ignore a lot of what they know.

To test whether information matters, Mookerjhee and Sopher (1994) compared high- and low-information conditions. They found some differences, casting doubt on reinforcement models, but they used matching pennies games, which have little power to distinguish learning theories when simple statistics are used.

A more thorough test was conducted by Van Huyck, Battalio, and Rankin (2001a). They use an order-statistic coordination game with five players. A player's payoff depends on her own choice and the median choice. Strategies s_i^j are numbers in the interval [0,1]; denote the vector of all other players' strategies by s_{-i}; and $M(s(t))$ is the median of all five players' strategies. Player i's payoff is then

$$\pi_i(s_i^j, s_{-i}) = .5 - |s_i^j - \omega \cdot M(s) \cdot (1 - M(s))|. \tag{6.2.7}$$

In their game with $\omega = 2.44$, denoted G(2.44), the maximum value of $\omega \cdot M(\cdot) \cdot (1 - M(\cdot))$ is 0.61, so any number above 0.61 is dominated. The Nash equilibria are solutions to the equation $x = 2.44 \cdot x \cdot (1 - x)$, which are 0 and 0.59.

This can be seen graphically in Figure 6.1 phase diagrams, which show how a median $M(t+1)$ (on the y-axis) depends on the previous median $M(t)$ (on the x-axis). The inverted U-shaped line plots the best-response function $M(t) = 2.44M(t)(1 - M(t))$. Equilibria are where the best-response function and equilibrium condition $M(t) = M(t+1)$ intersect (i.e., 0 and 0.59).

Van Huyck et al. (VHBR) ran four sessions with G(2.44). Subjects knew they were playing seventy-five periods with four others, and knew their own history of choices and payoffs, but did *not* know anything at all about the payoff function. The path of actual medians in the four experimental sessions are shown in the Figure 6.1 phase diagrams. Subjects zig-zag around the interior equilibrium at 0.59 for a while, then settle down after about twenty periods.

Van Huyck et al. ask whether the actual dynamic path in Figure 6.1 can be explained by the Cross (1973) reinforcement rule. The Cross rule modifies choice probabilities according to

$$P(s_i^j)(t+1) = P(s_i^j)(t) + \alpha \cdot r[\pi_i(s_i^j, s_{-i}(t))] \cdot (1 - P(s_i^j(t)) \qquad s_i^j = s_i(t) \tag{6.2.8}$$

$$P(s_i^j)(t+1) = P(s_i^j)(t) - \alpha \cdot r[\pi_i(s_i^j, s_{-i}(t))] \cdot P(s_i^j(t)) \qquad s_i^j \neq s_i(t), \tag{6.2.9}$$

where $r(\pi)$ is the reinforcement from payoff π.[12]

Figure 6.2 shows a simulated path of the Cross dynamic for seventy-five periods for $\alpha = 0.05$ and uniform initial conditions. The process converges much more slowly than the actual data, and often wanders into the dominated-strategy territory above 0.61 (which subjects rarely do). The Cross reinforcement rule clearly learns far too slowly to explain the behavior of subjects. The adjustment speed α can be increased but, as Van Huyck et al. note (2001a, p. 15), "when you speed up these algorithms they begin to get stuck in absorbing states that are not even close to being mutually

[12] They normalize reinforcements to be 0 and 1 for the lowest and highest payoffs, and try adjustment speed parameters $\alpha = 0.01$ or 0.05.

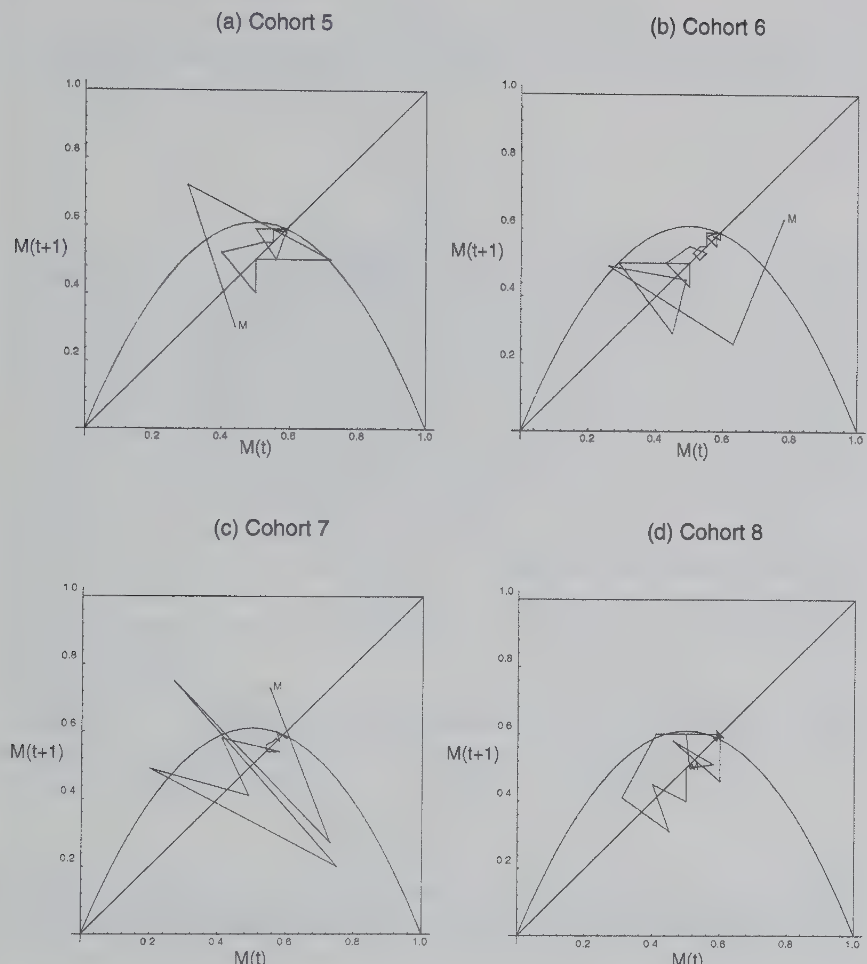


Figure 6.1. *Phase diagrams of actual period-by-period medians in Van Huyck, Battalio, and Rankin's order-statistic games. Source: Van Huyck, Battalio, and Rankin (2001a).*

consistent, something our subjects don't do, and if one continually smears probability to prevent this then the process doesn't converge, which is also something our subjects don't do." However, Sarin and Vahid (2000) show that spilling reinforcement to a chosen strategy 6–12 neighboring strategies, and decaying lagged attractions rapidly, improves the fit of the model substantially.

Van Huyck et al. compare these sessions to complete-information sessions in which subjects knew the payoff function. Reinforcement learning predicts that the additional payoff function information should not matter

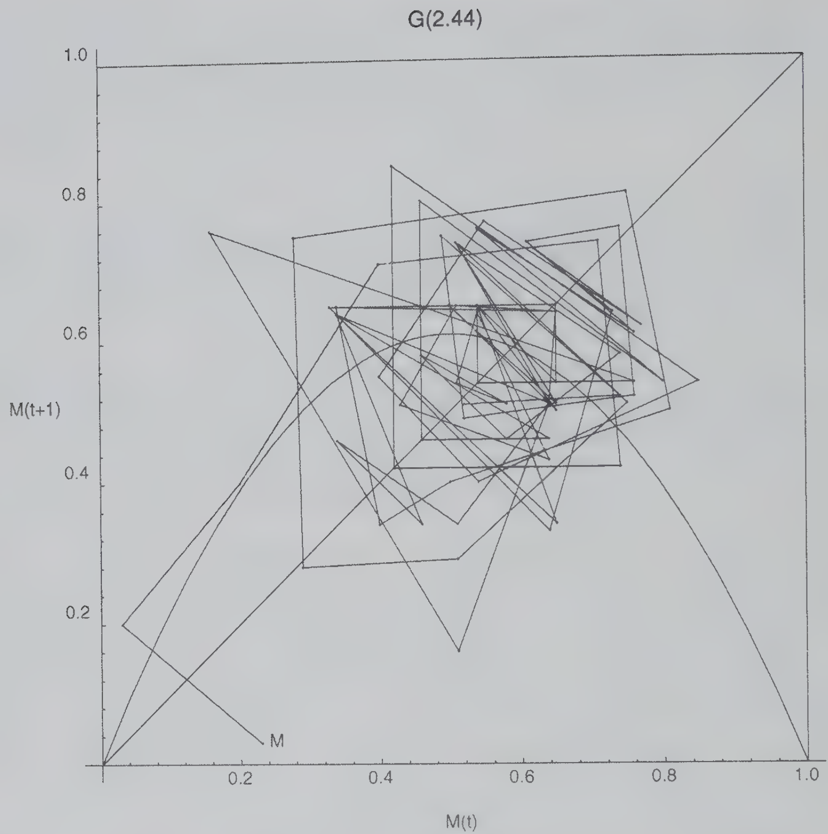


Figure 6.2. *Simulated paths of Cross (reinforcement) dynamics in Van Huyck, Battalio, and Rankin's games (G(2.44)). Source: Van Huyck, Battalio, and Rankin (2001a).*

but subjects converge much faster when they know the payoff function (in five periods rather than twenty).

Summary: Reinforcement learning models in which only chosen strategies are reinforced predict the direction of learning correctly, but are often too slow to match the pace of human learning (except in games with mixed equilibria, when there is little learning and many models improve slightly over equilibrium predictions or random guessing). Underestimation of learning can be directly traced to the fact that, in environments with full payoff information, subjects use that information. One can speed up the algorithms to approximate human learning by assuming that players reinforce all strategies according to unchosen payoffs, reinforce "winning" strategies, or reinforce a wide range of strategies that are "similar" to the one they chose. Models discussed later in this chapter move in the same direction, by expanding the range of how subjects learn.

6.3 Belief Learning

An early model of belief learning dates back to Cournot ([1838] 1960), who suggested players choose a best response to behavior observed in the previous period. Theories of "fictitious play" were proposed by Brown (1951) and Robinson (1951). These theories were initially proposed to compute Nash equilibrium algorithmically, providing a story about how mental simulation could lead to immediate equilibration in a kind of cognitive tâtonnement.

Early proofs showed that in some games, if fictitious play beliefs converged, they would converge on a Nash equilibrium. Then, in 1964, Shapley showed that, in a certain zero-sum game with three strategies, fictitious play cycles around strategies and does not converge to the unique mixed-strategy equilibrium. His counterexample took the wind out of the sails of the hope for a general guarantee that fictitious play would always converge to Nash equilibrium. Research on learning dynamics came to a screeching halt for about fifteen years.

In the 1980s, theorists dusted off fictitious play and reinterpreted it as a theory of how players might learn from periods of actual behavior rather than from mentally simulated trials (Fudenberg and Kreps, 1995).

6.3.1 Weighted Fictitious Play

Cheung and Friedman (1997) were the first to study belief learning thoroughly using data. They use a weighted form of fictitious play and a stochastic response function. Weighted fictitious play beliefs are defined by

$$E_i(s_{-i}^j)(t) = \frac{I(s_{-i}^j, s_{-i}(t)) + \sum_{k=1}^{t-1} \phi_i^k \cdot I(s_{-i}^j, s_{-i}(t-k))}{1 + \sum_{k=1}^{t-1} \phi_i^k} \quad t = 1, 2, \ldots \quad (6.3.1)$$

(i.e., observations t periods back are weighted by ϕ^{t-1}). Adding the weight ϕ is sensible because standard fictitious play ignores *when* another player made various choices in the past. Cournot best-response dynamics errs in the opposite direction by taking into account only what happened last time. Weighted fictitious play is a sensible compromise. When $\phi = 1$, weighted fictitious play is original fictitious play; when $\phi = 0$, it is Cournot.[13]

[13] A further generalization allows ϕ to vary with time (e.g., Crawford, 1995). For example, ϕ might grow over time if players recognize that other players are changing their strategies more frequently at the beginning of a game than at the end (e.g., Ho, Camerer, and Chong, 2002).

Table 6.7. Games used by Cheung and Friedman

Game	Matrix	
Hawk–dove	−2, −2	8, 0
	0, 8	4, 4
Coordination	5, 5	−1, 4
(stag hunt)	4, −1	1, 1
Buyer–seller	2, 2	0, 3
	3, −1	−1, 4
Battle of the sexes	1, 3	−1, −1
	−1, −1	2, 1

Source: Cheung and Friedman (1997).

Their games are shown in Table 6.7. They varied the matching and history protocols in interesting ways. In a (standard) random pairwise (RP) condition, players were randomly rematched in each period. In a "mean matching" (MM) condition, each player chose one strategy and earned the mean of their payoffs averaged over matches with every other player. In the (standard) no history (NH) condition, each player knew only their own personal history; in a history (H) condition players knew the distribution of *all* strategies chosen previously.

Parameters are estimated separately for each individual. There are many estimates in outlying categories, greater than 1 or below 0 (which implies parameters are not estimated too precisely). The median fictitious play weights ϕ are between 0.25 and 0.50, closer to the Cournot prediction of 0 than the fictitious play prediction of 1.

Table 6.8 classifies individuals into types according to the value of ϕ.[14] For example, Cournot types are those for whom the hypothesis that $\phi_i = 0$ can be accepted but $\phi_i = 1$ can be rejected. Adaptive types are those for whom both $\phi_i = 1$ and $\phi_i = 0$ can be rejected. Most estimates of ϕ tend either to include one extreme but not the other (Cournot and fictitious play), or to include both (uninformative). The type distribution is not significantly different across the four games. Cheung and Friedman note that, since players have more information when they know all players' histories, or are matched to everyone and paid the average, they should weight new

[14] The two "imprinting" types with $\hat{\phi} > 1$ (also known as "primacy effects" in cognitive psychology) are excluded.

Table 6.8. *Number of subjects classified by estimated fictitious play weight*

Game type	Fickle	Cournot	Adaptive	Fictitious play	Uninformative
Accept	$\phi_i < 0$	$\phi_i = 0$	$0 < \phi_i < 1$	$\phi_i = 1$	$\phi_i = 0,1$
Reject	$\phi_i = 0$	$\phi_i = 1$	$\phi_i = 0,1$	$\phi_i = 0$	
Hawk–dove	15	31	8	31	41
Coordination	5	25	6	18	22
Buyer–seller	4	37	9	17	25
Battle of the sexes	4	51	6	18	18

Source: Cheung and Friedman (1997).

information more heavily (higher ϕ_i) and respond more sensitively (higher λ_i) in those conditions. These predictions are correct.

6.3.2 General Belief Learning

Crawford (1995; and see Crawford and Broseta, 1998) proposes a belief-type approach to model how the interaction of dispersed initial beliefs and adaptive dynamics explain convergence patterns in order-statistic coordination games studied by Van Huyck, Battalio, and Beil (1990, 1991) (see Chapter 7). The model allows common changes in beliefs and player-specific shocks.

In these games, players choose a number from 1 to 7 and a player i's payoff depends on the number she chose, $s_i(t)$, and on an order statistic of all the numbers, $y(t)$. Payoffs are generally increasing in $y(t)$ and decreasing in deviation between $s_i(t)$ and $y(t)$. In Van Huyck et al. (1990) the order statistic was the median. In these games, players are penalized quadratically for choosing a number higher or lower than the median number, which induces a desire for conformity. In Van Huyck et al. (1991) the order statistic was the minimum. In these "weak-link" games, players were penalized if their choice was above the minimum, but everyone wanted the minimum to be high.

In the median game, there is some dispersion in initial choices, with frequent choices of 4–5, and a fast convergence to the initial median in a couple of periods. In minimum games with large groups, the dispersion in initial choices means that the minimum of those choices is usually low, and the large groups *always* converge to the minimum of 1, usually within a few periods.

Crawford's analysis exploits the fact that, since players are penalized for deviating from the order statistic, there is no need to distinguish between players' *beliefs* about the order statistic and their *choices*. Following the literature on adaptive control, he assumes that player i's initial choices $s_i(1)$ and adapted choices $s_i(t)$ are given by

$$s_i(1) = \alpha_0 + \zeta_{i0} \tag{6.3.2}$$

$$s_i(t) = \alpha_t + (1 - \phi_t)y_{t-1} + \phi_t s_i(t-1) + \zeta_{it} \quad (t = 1, 2, \ldots). \tag{6.3.3}$$

In each equation, an individual player's response depends on common parameters and a player-specific term ζ_{it}.

The intercepts α_t capture period-specific drift in choices, which is common across players. A lower value of the weight ϕ_t represents quicker responsiveness to experience. The terms ζ_{it} represent idiosyncratic player-specific initial beliefs and shocks to beliefs.[15]

Under special parametric restrictions, the model reduces to fictitious play or Cournot. But the full model is a generalization of these approaches, which allows time-varying history weights and idiosyncratic changes in beliefs (owing to bursts of insight or correlated shocks due to public events). Solving recursively, Crawford shows that for a given set of parameter values there is a unique path of choices and order statistics. If the weights $(1 - \phi_t)$ lie in an interval $(0, 1 - b)$ for some $b < 1$, and the summed common drift terms α_t and shock variances $\sigma^2(\zeta_{it})$ are finite, all players eventually choose the same number. Intuitively, this means a group will equilibrate (reaching a mutual best-response point) as long as they continually respond to experience and the drift and shock terms dampen over time.

These conditions imply convergence will occur, but *which* equilibrium the players converge to will depend stochastically on initial conditions and parameter values (i.e., even an observer knowing only the initial conditions and parameters cannot predict exactly what will happen). As Crawford emphasizes, this is a kind of proof that experimental observation is necessary: Theorizing alone cannot predict which paths will occur, so putting people in these games and observing what happens is necessary to understand the system fully.

Some econometrics is used to estimate the model's parameters. To conserve degrees of freedom, some innocuous restrictions are imposed. The estimate of α_0 is 4.75, which matches the central tendency in the data to choose 4 or 5 in the first period. The variance of the first-period player-specific shock is $\sigma^2(\epsilon_t) = 1.62$, implying a standard deviation of 1.27 of initial choices across players. The learning coefficient $\hat{\phi}$ is 0.42, which means that

[15] If the conditional variance $\sigma^2(\zeta_{it}|x_{it-1})$ is 0, then all players who chose the same number in the last period have the same idiosyncratic change in beliefs. This can be shown to imply, surprisingly (see Crawford's footnote 18), that the order statistic y_t must converge to its initial value y_0. Since this path-dependence is evident in median games, Crawford infers that the conditional variance of the player-specific shocks must be small.

players weight their previous observed median y_t slightly more (0.58) than their own previous choices (0.42).

In minimum-action games with large groups, the drift term α_t is constrained to be constant across periods t, but is allowed to be nonzero. It is estimated to be -0.27, reflecting the downward drift across periods evident in the data. The estimated inertia coefficient $\hat{\phi}$ is larger, 0.75, indicating that players are more inertial than in the median-action games.

Broseta (2000) adds an ARCH (autoregressive conditional heteroskedasticity) modification to the Crawford model. In ARCH, the unconditional shock variances in each period t, $\sigma^2(\zeta_{it})$, are positively autocorrelated over time. That is, $\sigma_t^2 = \kappa_{0t} + \kappa_1 \sigma_{t-1}^2$, where κ_{0t} is a period-specific component of unconditional variance and κ_1 captures the correlation between player-specific variances in adjacent periods. If $\kappa_1 > 0$, when player i has an idiosyncratic shock in a period that is large in absolute value (meaning that ζ_{it}^2 was large for i in period t), she is likely to have large shocks in subsequent periods (ζ_{it+1}^2 will be large too). Under mild conditions on the conditional variances (κ_{0t}s converge to 0 quickly enough, and κ_1 is not too close to 1), the ARCH process leads to convergence of players' choices in the limit, as in Crawford's setup.

The ARCH specification is like an omnibus correction for omitted variables which persistently influence players' forecasts. Allowing correlation of error variances across time will improve the overall fit, without requiring one to specify precisely what the omitted variable is. As Broseta (2000, p. 34) explains:

> To gain some intuition, suppose that for unknown cognitive reasons, subject k expects a high value of the order statistic and, accordingly, tends to play a high effort level at time t. We should then expect to observe, ex post, a large (and in this case, positive) prediction error ϵ_{kt}. After the value of y_t is publicly announced, all players update their beliefs. But if they do so sensibly . . . idiosyncratic differencs are likely to persist in the near future. In particular we would expect that, as he or she clings to "optimistic" beliefs, agents k will still expect a high value of y_{t+1}, which is then likely to result, ex post, in a large (and positive) value of ϵ_{kt+1}.

The ARCH model fits substantially better even though it is simpler. Broseta imposes some economizing restrictions. Most coefficients are similar to Crawford's. The ARCH persistence coefficient κ_1 is estimated to be 0.40 and 0.63 in median and minimum games, which is large.

Many other papers study belief models but are not discussed in detail here. Cooper, Garvin, and Kagel (1997a,b) apply belief learning models to signaling models of entry deterrence. Chapter 8 discusses their work and related work on belief learning in signaling games by Brandts and Holt (1993,

1994) and Anderson and Camerer (2000). In a study of the dominance-solvable "travelers' dilemma" (see Chapter 5), Capra et al. (1999) fit a fictitious play model, which goes in the right direction but underpredicts the magnitude of change by a factor of four. Sefton (1999) models learning in coordination games with fictitious play.

6.3.3 Learning Direction Theory

Learning direction theory settles for predicting only the *direction* in which choices will change, in low-information environments where players may know only the ideal direction of change. Selten often uses the example of an archer shooting an arrow toward a target. If the arrow misses to the left of the target, the archer knows to aim further to the right, but may know little more.

Although learning direction theory has never been fully specified, it can be interpreted as a relative of belief learning which combines elements of Cournot dynamics—moving toward the ex post best response—with habit or inertia. Learning direction theory was first used to analyze experimental data on finitely repeated prisoners' dilemmas (PDs) by Selten and Stöcker (1986). Their subjects played twenty-five sequences consisting of a ten-period repeated PD. They define a "cooperative play" of a sequence as one in which both players choose C for at least four periods, one player chooses D in at least one subsequent period, and both players choose D in all subsequent periods. Since most sequences are played this way, it is reasonable to characterize sequences by the first period in which defection occurs. Table 6.9 shows the average period in which deviations were intended to occur, averaged across blocks of three sequences for six experimental sessions.[16] Sessions vary—for example, defection occurs around periods 6–7 in groups I and IV, and in period 8 or later in others. Players also defect earlier and earlier across sequences.

In their learning direction model, each subject has an intended deviation period k. Players who intended to defect later than their opponent actually did (i.e., the opponent defected before them) lower their k by one unit with probability α. Players who defect at the same time as their opponent lower their k by one unit with probability β.[17]

[16] "Intended" defection is either a player's first D period, or their intended period from an analysis of written responses. Means greater than 10 occur because period 11 denotes cooperation through all ten periods.

[17] Presumably $\alpha > \beta$, because players learn more and react more strongly to a defection that precedes their defection than to a defection that occurs at the same time. Notice that this intuition uses a concept of forgone payoff in the background, but the amount of the loss from defecting later does not enter directly into a calculation about the switch rates α and β.

Table 6.9. Mean intended defection periods in ten-period prisoners' dilemma

	Session						
Sequences	I	II	III	IV	V	VI	Total
13–15	7.7	8.8	10.1	7.6	10.0	10.2	9.1
16–18	7.0	8.5	9.9	7.3	9.3	9.9	8.7
19–21	6.4	8.0	9.8	6.8	9.2	9.2	8.2
22–25	5.7	7.8	9.5	6.1	8.7	8.5	7.6

Source: Selten and Stöcker (1986).

If a player defects *before* her opponent she increases k by one unit with probability γ. In most cases (65 percent) players change their defection period as predicted by direction learning; in 35 percent of the cases players changed in the wrong direction. Counting the relative frequencies of changes gives the median estimates of α, β, and γ of 0.500, 0.135, and 0.225. The median estimate of the difference $\alpha + \beta - \gamma$ is 0.45, 78 percent of those individual-level estimates are positive, and steady-state calculations using these parameters imply that, in the long run, unraveling to immediate defection will occur.

Learning direction theory imposes a simple structure on learning and usually predicts a majority of directional changes in other studies. However, it is hard to be thoroughly impressed by direction learning because it makes such an unsurprising prediction. The same kind of movement toward best response is built more generally into theories such as belief or EWA learning. In those theories, when the reinforcement on forgone payoffs is high enough, players will shift choice probability toward best responses just as in learning direction theory.

Furthermore, define the ex post best response

$$b_i(t) = \operatorname{argmax}_{s_i^j} \pi_i(s_i^j, s_{-i}(t)).$$

Learning direction theory predicts only that players are likely to choose strategies between $s_i(t)$ and $b_i(t)$. But imagine a game in which some strategies in the interval $[s_i(t), b_i(t)]$ have very low payoffs. Learning direction theory does not predict anything about whether those low-payoff interior strategies will be played or not. Theories that are responsive to forgone payoffs (such as beliefs and EWA) predict precisely which strategies in the interval are more or less likely to be played. Learning direction predictions are therefore sharpened when more information is available.

6.3.4 Bayesian Learning

Cox, Shachat, and Walker (2001) did the first experimental test of Jordan's (1991) Bayesian learning model. In that model, players are uncertain what other players' payoffs are, but they have a commonly known prior and can learn over time, from actions of other players, which payoff matrix is being used. Jordan showed that the Bayesian learning process converges to a Nash equilibrium in finite games (and sometimes it refines the set of equilibria). Although Jordan's result is reassuring, it is tailored to explain only learning about what other players' *payoffs* are, rather than learning what other players *will do*. In experiments where payoffs are common knowledge, the Jordan model predicts immediate equilibration, which we rarely see. Because the model's prediction is wrong in these simpler cases, it is not the best general approach. Nonetheless, it makes interesting predictions in games with incomplete payoff information which are worth examining.

In the experiments of Cox et al., row players have payoffs given by one of four matrices with commonly known priors, and column players have payoffs given by one of two matrices. All matrices are shown in Table 6.10. Each true game is a combination of a row player payoff matrix (with priors shown in parentheses) and a column player matrix. Players know their own matrix but not the other player's matrix.

Along the Bayesian equilibrium path, players take actions but also learn about the other player's payoff matrix (and likely actions) over time. (Many theorists love this model because it characterizes "equilibrium learning"–that is, players are changing what they do, but are always perfectly anticipating what others will do given what they know.)

To illustrate, suppose the matrix combination drawn is C for row and B for column (denoted CB). Since the row player has a dominant strategy, she plays (B)ottom for sure and earns 2. Column knows there is a 3/8 chance the matrix is A, and row will play (T)op, and a 3/8 chance the matrix is C, and row will play B. In equilibrium, row also plays B when the matrix is B and T

Table 6.10. *Games used in experiments on Bayesian (Jordan) learning*

	Possible row player payoff matrices								Possible column player payoff matrices			
	A (3/8)		B (1/8)		C (3/8)		D (1/8)		B (1/2)		D (1/2)	
	L	R	L	R	L	R	L	R	L	R	L	R
T	1	2	0	3	1	0	3	0	0	2	3	1
B	0	0	2	2	2	2	1	1	3	2	0	1

Source: Cox, Shachat, and Walker (2001).

when the matrix is D (these inferences come from more delicate reasoning). If the column player anticipates these moves and their probabilities, she infers that the row moves will be T, B, B, T for row matrices A–D, so, given the priors, there is a 50–50 chance row will play either T or B. Given that guess and knowing her own payoffs are given by matrix B—in our example of the CB combination—column should play R and earn 2. Hence, the Jordan path predicts (B,R) play in period 1 when the matrix combination is CB.

Notice that if they play (B,R), column then earns 2 but also learns that the row player's matrix is either B or C. Then column infers row will keep playing B so she should switch to L and earn 3. The second-period prediction is therefore (B,L). Then there is nothing more to be learned and the players should play (B,L) (a Bayesian–Nash equilibrium) forever.

In the first period, as noted, both types of players with matrices B and D should choose B and T, or R and L, respectively. Row players make that predicted choice about 70 percent of the time and column players make it 60 percent of the time. The corresponding figures in the second period (conditional on period 1 history) are 54 percent and 69 percent. Players are therefore roughly on the Jordan path but there are many deviations.

Table 6.11 shows relative frequencies of each strategy pair and matrix pair, from the third period on (when the Jordan path is predicted to settle into equilibrium), averaging across matching protocols. Jordan predictions

Table 6.11. *Summary results in Bayesian learning experiment from period 3 on*

	Strategy pair			
Matrix pair	(T,L)	(T,R)	(B,L)	(B,R)
AB	0.03	**0.97**	0.00	0.01
AD	**0.91**	0.05	0.05	0.00
CB	0.00	0.00	**0.96**	0.04
CD	0.00	0.00	0.09	**0.91**
BB	0.13	*0.26*	**0.37**	0.23
DD	**0.70**	0.08	0.10	*0.12*
BD	0.14	0.23	0.32	0.31
DB	0.14	0.27	0.23	0.35
Mixed-strategy predictions (DB, BD)	0.11	0.22	0.22	0.44

Source: Cox, Shachat, and Walker (2001).
Note: Jordan predictions are in **bold**; Nash predictions that do not coincide with Jordan are in *italics*. Mixed-strategy predictions apply to games DB and BD, and are also Nash equilibria (but not Jordan) in DD and BB.

are in bold. When the row player has a dominant strategy, in matrix pairs (AB,AD) and (CB,CD), more than 90 percent of strategy pairs are at the Jordan and Nash prediction. In matrix pair DD the Jordan prediction of (T,L) is again rather accurate (70 percent) but in the matrix pair BB the predicted (B,L) is played only about a third of the time. In games DB and BD the Jordan and Nash predictions are mixed strategies in which (T,L) is most rare and (B,R) is most common, and the data do match the predicted rank of frequency across pairs (as in most mixed games described in Chapter 3).

The predictions of the Jordan model depend delicately on players making the right choices in the first couple of periods. Given this fragility, the results above are encouraging. At the same time, these games are only 2×2 and the row player sometimes has dominant strategies, so Bayesian learning is not computationally difficult. It should be easy to construct games that are only slightly more complicated in which the Jordan paths are very unlikely to emerge in the short run.

6.3.5 Measuring Beliefs Directly

Several studies in experimental economics have measured players' beliefs using incentive-compatible "scoring rules" to induce players to report thoughtfully.[18] In a pioneering study, McKelvey and Page (1990) elicited beliefs to test information aggregation. In 1988, Weigelt and I elicited beliefs to test refinement predictions about out-of-equilibrium beliefs in signaling and trust games (reported in Camerer, Ho, and Chong, 2002b). Karjalainen and I (1994) found elicited beliefs in a coordination game could be superadditive, reflecting ambiguity-aversion owing to strategic uncertainty.

Many years later, Nyarko and Schotter (2002) elicited beliefs to answer an obvious but unanswered question—do beliefs correspond to fictitious play? (The answer is no.) They used a 2×2 game with a unique mixed-strategy equilibrium (MSE), shown in Table 6.12. Subjects played sixty rounds in four sessions, crossing random- and fixed-matching protocols with and without belief elicitation. As in other mixed games (see Chapter 3), actual frequencies across all periods are between equal randomization and the MSE prediction.

Because beliefs are measured directly, one can fit stated beliefs to a weighted fictitious play model to see how well it explains beliefs. These fits

[18] For example, Nyarko and Schotter's subjects play either R(ed) or G(reen). If the player reports a belief r that the opponent will play R, they earn $0.10 - 0.10(1 - r)^2$ if R is actually played and $0.10 - 0.05(r^2 + (1 - r)^2)$ if G is played. It is easy to show that, if players are risk neutral and have a true belief b, then their expected payoff is maximized by reporting $b = r$ (e.g., Camerer, 1995).

Table 6.12. *Nyarko and Schotter's mixed-strategy game*

Row strategies	Column strategies Green	Red	MSE prediction	Relative frequency
G	6,2	3,5	0.40	0.46
R	3,5	5,3	0.60	0.54
MSE prediction	0.40	0.60		
Relative frequency	0.39	0.61		

Source: Nyarko and Schotter (2002).

yield ϕ coefficients close to 1[19] but stated beliefs and fitted beliefs are usually very far apart. A key problem is that fictitious play beliefs with ϕ close to 1 average all previous observations and settle down rapidly after about twenty periods. But stated beliefs vary wildly from period to period and do not converge. Players' actions are also more like best responses given stated beliefs, *not* like best responses given fictitious play beliefs. However, stated beliefs are slightly *worse* predictions of actual behavior by opponents than are fictitious play beliefs (i.e., subjects would have earned more money if they stated fictitious play beliefs instead of what they actually guessed).

The Nyarko and Schotter results show fictitious play is a bad model of stated beliefs in these games. Beliefs may be drawing on more information than is used in fictitious play, or may reflect time-variation of weights (see Camerer and Ho, 1999b) or sophisticated outguessing of how others are learning.

6.3.6 Population-Level Replicator Dynamics

Cheung and Friedman (1998) apply replicator dynamics to their hawk–dove and buyer–seller games and compare it with individual learning.[20] In replicator dynamics, the percentage of a population that uses a strategy increases proportionally with the relative payoff advantage of that strategy. The payoff advantage to strategy 1, which is strategy 1's expected payoff minus the population average payoff, is $(1 - S_t) \cdot R(S_t)$, where S_t is the proportion playing 1. The replicator dynamic is defined as $(S_{t+1} - S_t)/S_t = \beta \cdot (1 - S_t)R(s_t)$, where β is an adjustment speed parameter. Note that replicator dynamics predicts only how population-level averages change; it makes no predictions about how specific individuals will change what they do.

[19] The median $\hat{\phi}$ across subjects is 1.05 and the interquartile range is (0.98, 1.32). Inferring ϕ coefficients from choices yields a lower overall estimate of $\hat{\phi} = 0.61$ with much variability. Note that Feltovich (2000) also finds good fits from ϕ slightly greater than 1, which is odd and merits further investigation.

[20] This section benefited from discussion with Dan Friedman.

Table 6.13. Replicator dynamics coefficients

Sample	Estimate $\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	Mean absolute deviation Replicator	Weighted FP
Hawk–dove	−.08 (.01)	.50 (.04)	.39 (.06)	.119	.117
Buyer	.23 (.05)	.60 (.26)	.23 (.54)	.22	.16
Seller	−.07 (.51)	.49 (.03)	.48 (.14)	.19	.16

Source: Cheung and Friedman (1998).
Note: Standard errors in parentheses.

Cheung and Friedman test replicator dynamics by regressing percentage changes of the population frequency against the historical relative payoff advantage

$$\delta S_{t+1}/S_t = \alpha + \beta \cdot (1 - S_t)R(s_t) + \gamma \cdot MM \cdot (1 - S_t)R(S_t) + \epsilon_t. \quad (6.3.4)$$

The intercept α captures drift toward strategy 1 (which is difficult for standard replicator dynamics to explain). The variable MM is a dummy which equals 1 if the data came from the mean-matching treatment, so γ measures any variable *increase* in adjustment speed from mean-matching.

Table 6.13 shows the results and compares the accuracy of replicator dynamics with weighted fictitious play (FP). The estimates of β are within the bounds necessary to ensure stability. Positive estimates of $\hat{\gamma}$ show that adjustment speeds are larger in the mean-matching (MM) condition in which subjects have more information.

Although replicator dynamics is very popular in evolutionary game theory, there are three problems with it as a theory of population learning in the lab. First, it doesn't fit as well: The mean absolute deviations from actual population frequencies are simply higher than for the weighted fictitious play model. Second, there is no sensible reason why the intercepts $\hat{\alpha}$ should be nonzero—strictly speaking, nonzero α implies that, even when strategy 1 has no payoff advantage (e.g., at the mixed-strategy equilibrium), there is drift toward play of the first strategy (for Sellers) or away from it (in Hawk–dove and Buyer). Third, in the Buyer–Seller game the replicator dynamic is "area-preserving," which means that, if the population begins in some region of population frequencies around the equilibrium, there is no pressure for the population to move closer toward the equilibrium over time. However, actual behavior *does* move closer. These drawbacks mean it is hard to take replicator dynamics seriously, relative to individual-level models, as the best possible model of human learning in high-information conditions.

Summary: The idea that players learn in games by updating beliefs about what others will do is a natural model with a long history. Empirical papers by Crawford and Broseta, Cheung and Friedman, Cooper and Kagel, Capra et al., and Sefton concentrate on models in this class. Their work suggests that belief-based models generally improve on Nash equilibrium. Without comparing these models with others, however, it is hard to know whether they truly fit well or not; the next section shows that they usually fit worse than other models.

One study shows that Jordan's Bayesian learning model predicts well in a simple setting, but that model cannot explain learning in complete-information games and probably predicts poorly in more complex games. Nyarko and Schotter directly measure beliefs that subjects state in a 2×2 mixed-equilibrium game. Stated beliefs deviate persistently from fictitious play beliefs (even though stated beliefs are slightly less accurate guesses), which raises an important question of which models can characterize how stated beliefs change, if fictitious play cannot.

6.4 Imitation Learning

People often learn by imitating actions of others. Imitation is especially prevalent among animals and children. And imitation is often a good economizing heuristic because players need only to repeat observed strategies, rather than form beliefs and evaluate all strategies (see Schlag, 1999).

Hück, Normann, and Oechssler (1999) studied imitation in Cournot oligopoly. Four players simultaneously choose outputs $q_i(t)$ in the interval [0, 100]. Prices are set by an inverse demand function $P(t) = 100 - Q(t)$, where $Q(t) = \sum_{i=1}^{4} q_i(t)$ is total quantity. Marginal cost is 1 for all players, so an individual player's profit is $\pi_i(t) = q_i(t)(P(t) - 1)$. The four players choose quantities in each of forty periods. Hück et al. compare behavior in four conditions, designed to test learning theories that require use of different kinds of information. Table 6.14 shows the four main information conditions.

Table 6.14. *Information conditions and output predictions in Hück et al.*

	Information given to firm i		Predicted output $Q(t)$		
Condition	Information	Outputs, profits	Cournot	Imitate-the-best	Mean $Q(t)$ (std. dev.)
BEST	demand, cost	Q(t)	79.2	n/a	82.56 (2.5)
FULL	demand, cost	$q_j(t), \pi_j(t) \forall j$	79.2	99	91.60 (6.5)
IMIT	none	$q_j(t), \pi_j(t) \forall j$	n/a	99	138.85 (31.6)
NOIN	none	$\pi_i(t)$	n/a	n/a	93.55 (14.7)

Source: Hück, Normann, and Oechssler (1999).

Simple computations (such as those many readers of this book have done as economics homework) give three benchmark predictions for total output. The Cournot–Nash equilibrium is $Q(t) = 79.2$. The Walrasian or competitive outcome (where price equals marginal cost) is $Q(t) = 99$. The collusive outcome, which maximizes total industry profit, is $Q(t) = 49.5$. The focus of Hück et al. is on the learning dynamics across information conditions. They build in inertia by allowing firms to change their output only with probability 1/3 in each period. Inertia stabilizes Cournot best-reply dynamics.[21]

The information conditions distinguish the predictions of best-response and imitation learning. If players learn by Cournot dynamics, they can calculate best replies with BEST and FULL information, and will converge to total outputs of 79.2. Since they cannot calculate best replies in the NOIN, IMIT, and IMIT+ conditions (they know only quantities and profits, but not costs), the Cournot learning theory makes no prediction in those conditions.

Imitation dynamics predicts that in the FULL, IMIT, and IMIT+ conditions players will converge to the competitive output of 99, since they know the most successful producer's output. Since they do not know who to imitate in the BEST and NOIN conditions, imitation theories make no prediction in those cases.

In the BEST condition, outputs creep downward toward the Cournot–Nash prediction of 79. In FULL, outputs drift upward toward the Walrasian prediction of 99. Because FULL subjects know individual outputs and profits, and imitate-the-best dynamics predicts convergence to the Walrasian prediction, this is the first glimpse of evidence that imitation occurs. In IMIT and IMIT+ the fluctuations in output are very large. However, in IMIT+ there is a visible tendency to converge toward the Walrasian output of 99. In NOIN the outputs fluctuate wildly at first then drift up toward the Walrasian output of 99 as well.

Hück et al. characterize learning with an omnibus regression that relates changes in output from period to period, $q_i(t) - q_i(t-1)$, to changes predicted by the different theories. Denote subject i's best reply to period $t-1$ output by $r_i(t-1)$, the quantity of the highest-profit producer in period $t-1$ by $b(t-1)$, and the average quantity of other firms' outputs by $a_i(t-1)$. Their regression is

$$q_i(t) - q_i(t-1) = \beta_0 + \beta_1(r_i(t-1) - q_i(t-1)) + \beta_2(b(t-1) - q_i(t-1)) \\ + \beta_3(a(t-1) - q_i(t-1)) + \epsilon_{it}. \tag{6.4.1}$$

[21] Vega-Redondo (1997) proved that if there is inertia (and some "mutation" or trembles), if firms imitate the quantity of the firm that is most successful, then quantities converge almost always to the Walrasian outcome. The intuition is simple—if price is above (below) marginal cost, then the largest (smallest) producer earns the most. If firms imitate the most successful producer, they are led to marginal-cost competitive pricing.

Table 6.15. *Quantity-change regression coefficients in Hück et al.*

Variable	Coefficient	Coefficient estimates BEST	FULL	IMIT	IMIT+
Best-reply	β_1	.430	.366	—	—
		(.038)	(.044)	—	—
Imitate-the-best	β_2	—	.110	.465	.435
		—	(.038)	(.046)	(.040)
Imitate-the-average	β_3	.340	.344	.151	.273
		(.038)	(.038)	(.048)	(.047)
	R^2	.410	.507	.356	.439
	N	610	631	620	533

Source: Hück, Normann, and Oechssler (1999).
Note: Standard errors in parentheses.

The coefficients $\beta_1, \beta_2, \beta_3$ measure the extent to which quantities change in the direction predicted by best-reply, imitate-the-best, and imitate-the-average learning rules. Results pooled across subjects are summarized in Table 6.15.

All coefficients are highly significant. In the FULL condition, the best-reply effect (.366) is a lot stronger than imitating the best (.110). But imitating the best looms large in the IMIT and IMIT+ conditions where best replies can't be calculated.

Bosch-Domenech and Vriend (in press) also studied imitation in two- and three-firm (duopoly and triopoly) quantity-setting experiments. Collusive, Cournot–Nash, and competitive Walrasian outcomes are shown in Table 6.16.[22] Players chose quantities from 8 to 32 simultaneously, for twenty-two periods. Like Hück et al., they varied information conditions to see when imitation was most common. In all conditions players learned the history of output and profit for all firms. In their "easy table" condition, players also saw a profit table showing possible profits for all combinations of outputs. In the "hard table" condition they received the same information in an opaque form—"an inconveniently arranged enumeration of the market prices associated with all output levels and possible cost levels." In the IMIT+ condition (which they call "hardest") subjects were told only that prices were lower when aggregate output was higher.

Recall that imitation of the most successful firm will lead subjects, in theory, to the competitive Walrasian outcome since the most successful firm

[22] In their experiments the inverse demand curves were $P(Q) = 414 - 4Q$ and $P(Q) = 530 - 4Q$ and cost functions were $C(q) = 174q - 146$ and $C(q) = 174q - 266$, respectively.

Table 6.16. Final outputs in duopoly and triopoly experiments

Actual output statistic	Information condition			Predictions		
	Easy table	Hard table	IMIT+	Collusive	Cournot–Nash	Walrasian
Two-firm duopoly						
Mean	18.2[a]	23.4[b]	22.4[b]	15	20	30
Median	20	24	20			
Mode (%)	15 (39%)	20,32 (17%)	15,30 (17%)			
Three-firm triopoly						
Mean	23.7[a]	24.3[ab]	26.4[b]	15	22	30
Median	23	24	27			
Mode (%)	23 (28%)	18,20,25 (11%)	24 (17%)			

Source: Bosch-Domenech and Vriend (in press).
Note: Identical letter superscripts denote outputs that are not significantly different at $p < .05$ (one-sided Wilcoxon test).

has the highest output. Bosch-Domenech and Vriend theorize that, since computing best responses (under time pressure of 1 minute per round) is more difficult in the hard-table condition, and essentially impossible in the IMIT+ condition, players will imitate the most successful player and gravitate toward higher (Walrasian) outputs more often in those conditions.

Outputs in the last two periods, shown in Table 6.16, show only weak support for that hypothesis. As it becomes harder for subjects to compute best responses—moving from the easy table to the hard table to IMIT+—outputs do rise slightly, but they are always significantly below the Walrasian output of 30.

Summary: Behavior in quantity-setting experiments shows both best-reply learning and imitation learning. These results and Stahl (2001), discussed below, suggest imitation should be taken seriously as an empirical source of learning. However, imitation may also be a heuristic shortcut to more general types of learning.

6.5 Comparative Studies

Many recent studies compare models in different types of "horse races." These comparisons are much more informative than the studies described earlier, because it is quite possible that some adaptive rule is an approximation to a similar or more general rule, and its weaknesses are revealed only by comparison.

6.5.1 Comparing Belief Models

Boylan and El-Gamal (1993) use a Bayesian procedure to draw inferences about the relative accuracy of Cournot and fictitious play. Their procedure starts with prior probabilities that each of the two theories are true. Then simulations are used to generate likelihoods for different observations. Using the priors, the simulated likelihoods, and the actual data, Bayes' rule can be used to infer the posterior probability that each theory is true. They apply this procedure to experiments on dominance-solvable games by Knott and Miller (1987) and coordination games by Cooper et al. (1990). Cournot is much worse on a dominance-solvable game and about equally good on a coordination game, so it is much worse overall.

A huge advantage of the Bayesian approach is the ability to naturally integrate results from different experiments. For example, if one theory is much better on one data set and much worse on another, by multiplying the likelihood ratios together (a procedure that assumes experiments are independent, which is sometimes unlikely) it is often possible to declare a clear overall winner rather than simply concluding that one study favored each theory.

6.5.2 Comparing Belief and Reinforcement Models

Several studies have compared belief and reinforcement learning.

Ho and Weigelt (1996) studied behavior in two-player coordination games with multiple pure-strategy equilibria (see Chapter 7). They compared four models of learning. One model is a simple form of reinforcement ("vindication") which essentially assumes very concave utility for payoffs. The other models are Cournot, reinforcement, and fictitious play, all of which are a modest improvement over a no-learning benchmark. Fictitious play fits best.

A thorough early study was done by Mookerjhee and Sopher (1997). They compared fictitious play, Cournot, and three types of reinforcement in two constant-sum games. They used games with four or six strategies, which are much better for distinguishing different theories than games with only two strategies (see Mookerjhee and Sopher, 1994; Salmon, 2001). Their games are shown in Chapter 3.

Their model is very general. Each strategy s_i^j has a score that is a weighted linear combination of the attractions $A_i^j(t-1)$ of *each* of player i's m_i strategies. Strategy s_i^j's score in period t is $S_j(t) = \sum_{k=1}^{m_i} \alpha_{jk} \cdot A_i^k(t-1)$. The probability that each strategy is chosen is a logit $P(s_i^j)(t) = e^{S_j(t)} / \sum_{k=1}^{m_i} e^{S_k(t)}$. This specification allows the attraction of one strategy k to affect the score of another strategy i, through the coefficients α_{ik}. Cross-effects occur if the attraction of one strategy affects probabilities of choosing other strategies *differently*; for example, if some strategies are regarded as more similar and treated as close strategic substitutes. Mookerjhee and Sopher use logit regression to estimate the probabilistic response model under six different hypotheses about how attractions are determined. Three hypotheses assume some kind of reinforcement learning and three correspond to types of belief learning (Cournot and ficititious play with various averaging i). The Cournot model fits poorly. Reinforcement fits slightly better than the belief-based model. Most individual cross-effect terms are insignificant, but the restriction that they are all 0 is often rejected at the 1 percent level. Overall, the Mookerjhee–Sopher findings are a messy win for reinforcement models.

Tang's dissertation (1996) reports experiments on three games with mixed-strategy equilibria. He specifies and estimates a dizzying variety of learning models using those data. I will not discuss the learning model estimation at length (the data are reported in Chapter 3) because it is complicated, the results are not very conclusive, and the statistic (MSD) used does not allow statistical inferences about which models are most accurate. However, variants of fictitious play do substantially worse than reinforcement models. The only models competitive with reinforcement are rule learning, in which players shift weight across different rules. None of the models is much better than just using the observed frequency of choices.

Battalio, Samuelson, and Van Huyck (2001) compared three stag hunt games with different best-response properties (see Chapter 7 for more on coordination games). The games are shown in Table 6.17. All three

Table 6.17. *Stag hunt games of Battalio et al.*

	Player 2 choices	
Player 1 choices	X	Y
Game 2R		
X	45,45	0,35
Y	35,0	40,40
Game R		
X	45,45	0,40
Y	40,0	20,20
Game .6R		
X	45,45	0,42
Y	42,0	12,12

Source: Battalio, Samuelson, and Van Huyck (2001).