1
Introduction

GAME THEORY IS ABOUT WHAT HAPPENS when people—or genes, or nations—interact. Here are some examples: Tennis players deciding whether to serve to the left or right side of the court; the only bakery in town offering a discounted price on pastries just before it closes; employees deciding how hard to work when the boss is away; an Arab rug seller deciding how quickly to lower his price when haggling with a tourist; rival drug firms investing in a race to reach patent; an e-commerce auction company learning which features to add to its website by trial and error; real estate developers guessing when a downtrodden urban neighborhood will spring back to life; San Francisco commuters deciding which route to work will be quickest when the Bay Bridge is closed; Lamelara men in Indonesia deciding whether to join the day's whale hunt, and how to divide the whale if they catch one; airline workers hustling to get a plane away from the gate on time; MBAs deciding what their degree will signal to prospective employers (and whether quitting after the first year of their two-year program to join a dot-com startup signals guts or stupidity); a man framing a memento from when he first met his wife, as a gift on their first official date a year later (they're happily married now!); and people bidding for art or oil leases, or for knick-knacks on eBay. These examples illustrate, respectively, ultimatum games (bakery, Chapter 2), gift exchange (employees, Chapter 2), mixed equilibrium (tennis, Chapter 3), Tunisian bazaar bargaining (rug seller, Chapter 4), patent race games (patents, Chapter 5), learning (e-commerce, Chapter 6), stag hunt games (whalers, Chapter 7), weak-link games (airlines, Chapter 7), order-statistic games (developers, Chapter 7), signaling (MBAs and romance, Chapter 8), auctions (bidding, Chapter 9).

In all of these situations, a person (or firm) must anticipate what others will do and what others will infer from the person's own actions. A game is a mathematical x-ray of the crucial features of these situations. A game consists of the "strategies" each of several "players" have, with precise rules for the order in which players choose strategies, the information they have when they choose, and how they rate the desirability (or "utility") of resulting outcomes. An appendix to this chapter describes the basic mathematics of game theory and gives some references for further reading.

Game theory has a very clear paternity. Many of its main features were introduced by von Neumann and Morgenstern in 1944 (following earlier work in the 1920s by von Neumann, Borel, and Zermelo). A few years later, John Nash proposed a "solution" to the problem of how rational players would play, now called Nash equilibrium. Nash's idea, based on the idea of equilibrium in a physical system, was that players would adjust their strategies until no player could benefit from changing. All players are then choosing strategies that are best (utility-maximizing) responses to all the other players' strategies. Important steps in the 1960s were the realization that behavior in repeated sequences of one-shot games could differ substantially from behavior in one-shot games, and theories in which a player can have private information about her values (or "type"), provided all players know the probabilities of what those types might be. In 1994, Nash, John Harsanyi, and Reinhard Selten (an active experimenter) shared the Nobel Prize in Economic Science for their pathbreaking contributions.

In the past fifty years, game theory has gradually become a standard language in economics and is increasingly used in other social sciences (and in biology). In economics, game theory is used to analyze behavior of firms that worry about what their competitors will do.[1] Game theory is also good for understanding how workers behave in firms (such as the reaction of CEOs or salespeople to incentive contracts), the spread of social conventions such as language and fashion, and which genes or cultural practices will spread.

The power of game theory is its generality and mathematical precision. The same basic ideas are used to analyze *all* the games—tennis, bargaining for rugs, romance, whale-hunting—described in the first paragraph of this chapter. Game theory is also boldly precise. Suppose an Arab rug seller can always buy more rugs cheaply, an interested tourist values the rugs at somewhere between $10 and $1000, and the seller has a good idea of how

[1] Game theory fills the conceptual gap between a single monopoly, which need not worry about what other firms and consumers will do because it has monopoly power, and "perfect competition," in which no firm is big enough for competitors to worry about. Game theory is used to study the intermediate case, "oligopoly," in which there are few enough firms that each company should anticipate what the others will do.

impatient the tourist is but isn't sure how much the tourist likes a particular rug. Then game theory tells you *exactly* what price the seller should start out at, and *exactly* how quickly he should cut the price as the tourist hems and haws. In experimental re-creations of this kind of rug-selling, the theory is half-right and half-wrong: it's wrong about the opening prices sellers state, but the rate at which experimental sellers drop their prices over time is amazingly close to the rate that game theory predicts (see Chapter 4).

It is important to distinguish *games* from game *theory*. Games are a taxonomy of strategic situations, a rough equivalent for social science of the periodic table of elements in chemistry. Analytical game *theory* is a mathematical derivation of what players with different cognitive capabilities are likely to do in games.[2] Game theory is often highly mathematical (which has limited its spread outside economics) and is usually based on introspection and guesses rather than careful observation of how people actually play in games. This book aims to correct the imbalance of theory and facts by describing hundreds of experiments in which people interact strategically. The results are used to create behavioral game theory. Behavioral game theory is about what players *actually* do. It expands analytical theory by adding emotion, mistakes, limited foresight, doubts about how smart others are, and learning to analytical game theory (Colman, in press, gives a more philosophical perspective). Behavioral game theory is one branch of behavioral economics, an approach to economics which uses psychological regularity to suggest ways to weaken rationality assumptions and extend theory (see Camerer and Loewenstein, 2003).

Because the language of game theory is both rich and crisp, it could unify many parts of social science. For example, trust is studied by social psychologists, sociologists, philosophers, economists interested in economic development, and others. But what *is* trust? This slippery concept can be precisely defined in a game: Would you lend money to somebody who doesn't have to pay you back, but might feel morally obliged to do so? If you would, you trust her. If she pays you back, she is trustworthy. This definition gives a way to measure trust, and has been used in experiments in many places (including Bulgaria, South Africa, and Kenya; see Chapter 3).

The spread of game theory outside of economics has suffered, I believe, from the misconception that you need to know a lot of fancy math to apply it, and from the fact that most predictions of analytical game theory are not well grounded in observation. The need for empirical regularity to inform

[2] To be precise, this book is only about "noncooperative" game theory—that is, when players cannot make binding agreements about what to do, so they must guess what others will do. Cooperative game theory is a complementary branch of game theory which deals with how players divide the spoils after they have made binding agreements.

game theory has been recognized many times. In the opening pages of their seminal book, von Neumann and Morgenstern (1944, p. 4) wrote:

> the empirical background of economic science is definitely inadequate. Our knowledge of the relevant facts of economics is incomparably smaller than that commanded in physics at the time when mathematization of that subject was achieved. . . . It would have been absurd in physics to expect Kepler and Newton without Tycho Brahe—and there is no reason to hope for an easier development in economics.

This book is focused on experiments as empirical background. Game theory has also been tested using data that naturally occur in field settings (particularly in clearly structured situations such as auctions). But experimental control is particularly useful because game theory predictions often depend sensitively on the choices players have, how they value outcomes, what they know, the order in which they move, and so forth. As Crawford (1997, p. 207) explains:

> Behavior in games is notoriously sensitive to details of the environment, so that strategic models carry a heavy informational burden, which is often compounded in the field by an inability to observe all relevant variables. Important advances in experimental technique over the past three decades allow a control that often gives experiments a decisive advantage in identifying the relationship between behavior and environment. . . . For many questions, [experimental data are] the most important source of empirical information we have, and [they are] unlikely to be less reliable than casual empiricism or introspection.

Of course, it is important to ask how well the results of experiments with (mostly) college students playing for a couple of hours for modest financial stakes generalize to workers in firms, companies creating corporate strategy, diplomats negotiating, and so forth. But these doubts about generalizability are a demand for more elaborate experiments, not a dismissal of the experimental method per se. Experimenters *have* studied a few dimensions of generalizability—particularly the effects of playing for more money, which are usually small. But more ambitious experiments with teams of players, complex environments, communication, and overlapping generations[3] would enhance generalizability further, and people should do more of them.

[3] See Schotter and Sopher (2000).

1.1 What Is Game Theory Good For?

Is game theory meant to predict what people do, to give them advice, or what? The theorist's answer is that game theory is none of the above—it is simply "analytical," a body of answers to mathematical questions about what players with various degrees of rationality will do. If people don't play the way theory says, their behavior has not proved the mathematics wrong, any more than finding that cashiers sometimes give the wrong change disproves arithmetic.

In practice, however, the tools of analytical game theory *are* used to predict, and also to explain (or "postdict"[4]) and prescribe. Auctions are a good example of all three uses of game theory. Based on precise assumptions about the rules of the auction and the way in which bidders value an object, such as an oil lease or a painting, auction theory then derives how much rational bidders will pay.

Theory can help explain why some types of auction are more common than others. For example, in "second-price" or Vickrey auctions the high bidder buys the object being auctioned at a price equal to the *second*-highest bid. Under some conditions these auctions should, in theory, raise more revenue for sellers than traditional first-price auctions in which the high bidder pays what she bid. But second-price auctions are rare (see Lucking-Reilly, 2000). Why? Game theory offers an explanation: Since the high bidder pays a price other than what she bid in a second-price auction, such auctions are vulnerable to manipulation by the seller (who can sneak in an artificial bid to force the high bidder to pay more).

How well does auction theory predict? Tests with field data are problematic: Because bidders' valuations are usually hidden, it is difficult to tell whether they are bidding optimally, although some predictions can be tested. Fortunately, there are many careful experiments (see Kagel, 1995; Kagel and Levin, in press). The results of these experiments are mixed. In private-value auctions in which each player has her own personal value for the object (and doesn't care how much others value it), people bid remarkably close to the amounts they are predicted to, even when the function mapping values into bids is nonlinear and counterintuitive.[5]

In common-value auctions the value of the object is essentially the same for everyone, but is uncertain. Bidding for leases on oil tracts is an example—different oil companies would all value the oil in the same way but aren't sure how much oil is there. In these auctions players who are most optimistic about the value of the object tend to bid the highest and win.

[4] In some domains of social science, these kinds of game-theoretic "stories" about how an institution or event unfolded are called "analytical narratives" and are proving increasingly popular (Bates et al., 1998).

[5] See Chen and Plott (1998) and the sealed-bid mechanism results in Chapter 4.

The problem is that, if you win, it means you were much more optimistic than any other bidder and probably paid more than the object is worth, a possibility called the "winner's curse." Analytical game theory assumes rational bidders will anticipate the winner's curse and bid very conservatively to avoid it. Experiments show that players do not anticipate the winner's curse, so winning bidders generally pay more than they should.

Perhaps the most important modern use of auction theory is to prescribe how to bid in an auction, or how to design an auction. The shining triumphs of modern auction theory are recent auctions of airwaves to telecommunications companies. In several auctions in different countries, regulatory agencies decided to put airwave spectrum up for auction. An auction raises government revenue and, ideally, ensures that a public resource ends up in the hands of the firms that are best able to create value from it. In most countries, the auctions were designed in collaborations among theorists and experimental "testbedding" that helped detect unanticipated weaknesses in proposed designs (like using a wind tunnel to test the design of an airplane wing, or a "tow-tank" pool to see which ship designs sink and which float). The designs that emerged were not exactly copied from books on auction theory. Instead, theorists spent a lot of time pointing out how motivated bidders could exploit loopholes in designs proposed by lawyers and regulators, and using the results of testbedding to improve designs. Auction designers opted for a design that gave bidders a chance to learn from potential mistakes and from watching others, rather than a simpler "sealed-bid" design in which bidders simply mail in bids and the Federal Communications Commission opens the envelopes and announces the highest ones. One of the most powerful and surprising ideas in auction theory—"revenue equivalence"—is that some types of auctions will, in theory, raise the same amount of revenue as other auctions that are quite different in structure. (For example, an "English" auction, in which prices are raised slowly until only one bidder remains, is revenue-equivalent to a sealed-bid "Vickrey" auction, in which the highest bidder pays what the second-highest bidder bid.) But when it came to designing an auction that actual companies would participate in with billions of dollars on the line, the auction designers were not willing to bet that *behavior* would actually be equivalent in different types of auctions, despite what theory predicted. Their design choices reflect an *implicit* theory of actual behavior in games that is probably closer to the ideas in this book than to standard theory based on unlimited mutual rationality. Notice that, in this process of design and prescription, guessing accurately how players will actually behave—good prediction—is crucial.[6]

[6] Howard Raiffa pointed this out many times, calling game theory "asymmetrically normative."

Even if game theory is not always accurate, descriptive failure is prescriptive opportunity. Just as evangelists preach *because* people routinely violate moral codes, the fact that players violate game theory provides a chance to give helpful advice. Simply mapping social situations into types of games is extremely useful because it tells people what to look out for. In their popular book for business managers, *Co-opetition*, Brandenburger and Nalebuff (1996) draw attention to the barest bones of a game—players, information, actions, and outcomes. Both are brilliant theorists who *could* have written a more theoretical book. They chose not to because teaching MBAs and working with managers convinced them that teaching the basic elements of game theory is more helpful.

Game theory is often used to prescribe in a subtler way. Sometimes game theory is used to figure out what it is likely to happen in a strategic interaction, so a person or company can then try to change the game to their advantage. (This is a kind of engineering approach too, since it asks how to improve an existing situation.)

1.2 Three Examples

This chapter illustrates the basics of behavioral game theory and the experimental approach with three examples (which are discussed in more detail in later chapters): ultimatum bargaining, "continental divide" coordination games, and "beauty contest" guessing games. Experiments using these games show how behavioral game theory can explain what people do more accurately by extending analytical game theory to include how players feel about the payoffs other players receive, limited strategic thinking, and learning.

The three games use a recipe underlying most of the experiments reported in this book: pick a game for which standard game theory makes a bold prediction or a vague prediction that can be sharpened. Simple games are particularly useful because only one or two basic principles are needed to make a prediction. If the prediction is wrong, we know which principles are at fault, and the results usually suggest an alternative principle that predicts better.

In the experiments, games are usually posed in abstract terms because game theory rarely specifies how adding realistic details will affect behavior. Subjects make a simple choice, and know how their choices and the choices of other subjects combine to determine monetary payoffs.[7] Subjects are

[7] These design choices bet heavily on the cognitive presumption that people are using generic principles of strategic thinking which transcend idiosyncratic differences in verbal descriptions of games. If choices are domain specific then the basic enterprise this book describes is incomplete; varying game labels to evoke

actually rewarded based on their performance because we are interested in extrapolating the results to naturally occurring games in which players have substantial financial incentives. The games are usually repeated because we are interested in equilibration and learning over time. An appendix to this chapter describes some key design choices experimenters make, and why they matter.

1.2.1 Example 1: Ultimatum Bargaining

I once took a cruise with some friends and a photographer took our picture, unsolicited, as we boarded the boat. When we disembarked hours later, the photographer tried to sell us the picture for $5 and refused to negotiate. (His refusal was credible because several other groups stood around deciding whether to buy their pictures, also for $5. If he caved in and cut the price, it would be evident to all others and he would lose a lot more than the discount to us since he would have to offer the discount to everyone.) Being good game theorists, we balked at the price and pointed out that the picture was worthless to him (one cheapskate offered $1). He rejected our insulting offer and refused to back down.

The game we played with the photographer was an "ultimatum game," which is the simplest kind of bargaining. In an ultimatum game there is some gain from exchange and one player makes a take-it-or-leave-it offer of how to divide that gain. Our picture presumably had no value to him and was valuable to us (worth more than $5 in sentimental value). A price is simply proposing a way to divide the gains from exchange between our true reservation price and his cost. His offer to sell for $5 was an ultimatum offer because he refused to negotiate.

In laboratory ultimatum games like this, two players, a Proposer and a Responder, bargain over some amount, say $10 (the sum used in many experiments). The $10 represents the value of the gain to exchange (or "surplus") that would be lost if the trade wasn't made. The Proposer offers x to the Responder, leaving herself $\$10 - x$. The Responder can either take the offer—then the Responder gets x and the Proposer gets $\$10 - x$—or reject it and both get nothing.

Because the ultimatum game is so simple, it is *not* a good model of the protracted process of most naturally occurring bargaining (and isn't intended to be). It *is* the right model of what happened to us after the cruise,

domain-specific reasoning is the next step. The study by Cooper et al. (1999) of ratchet effects in productivity games using Chinese factory managers—who face such effects in planned economies—is a good example (see Chapter 8).

and what happens in the waning minutes before a labor strike is called, or on the courthouse steps before a lawsuit goes to trial. It is a model of the last step in much bargaining, and hence is a building block for modeling more complicated situations (see Chapter 4).

Simple games test game-theoretic principles in the clearest possible way. Ultimatum games, and related games, also are useful for measuring how people feel about the allocations of money between themselves and others.

The analytical game theory approach to ultimatum bargaining is this: First assume players are "self-interested"; that is, they care about earning the most money for themselves. If players are self-interested, the Responder will accept the smallest money amount offered, say $0.25. If the Proposer anticipates this, and wants to get the most she can for herself, she will offer $0.25 and keep $9.75. In formal terms, offering $0.25 (and accepting any positive amount) is the "subgame perfect equilibrium".[8] By going first, the Proposer has all the bargaining power and, in theory, can exploit it because a self-interested Responder will take whatever she can get.

To many people, the lopsided distribution of the $10 predicted by analytical game theory (with self-interest) seems unfair. Because the allocation is considered unfair, the way people actually bargain shows whether people are willing to take costly actions that express their concerns for fairness. In the cruise-picture example, offering $1 instead of the $5 price the photographer offered added $4 to our surplus and subtracted $4 from his. If he thought this was unfair to him, he could reject it and earn nothing (even though everyone suffers—he earns no money and we don't get a picture we would like to own). The lab experiments simulate this simple game. Will Responders put their money where their mouths are and reject offers that seem unfair? If so, will Proposers anticipate this and make fair offers, or stubbornly make unfair offers?

In dozens of experiments conducted in several different countries, Proposers offer $4 or $5 out of $10 on average, and offers do not vary much. Offers of $2 or less are rejected about half the time. The Responders think much less than half is unfair and are willing to reject such small offers, to punish the Proposer who behaved so unfairly. Figure 1.1 shows data from a study by Hoffman, McCabe, and Smith (1996a). The x-axis shows the amount being offered to the Responder, and the y-axis shows the relative frequency of offers of different amounts. The dark part of each frequency bar is the number of offers that were rejected. Most offers are close to half

[8] Note also that every offer is a "Nash equilibrium" or mutual best-response pattern because x is the optimal offer if the Proposer thinks the Responder will reject any other offer. (This belief may be wrong but, if the Proposer believes it, she will never take an action that disconfirms her belief, so the wrong belief can be part of a Nash equilibrium.)

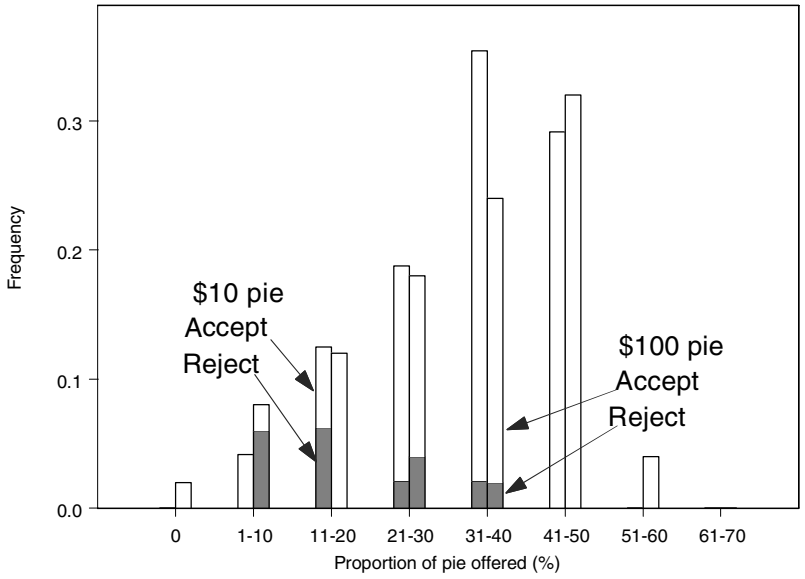


Figure 1.1. *Offers and rejections in high- and low-stakes ultimatum games. Source: Based on data from Hoffman, McCabe, and Smith (1996a).*

and low offers are often rejected. Figure 1.1 also shows that the same pattern of results occurs when stakes were multiplied by ten and Arizona students bargained over $100. (A couple of subjects rejected $30 offers!) The same basic result has been replicated with a $400 stake (List and Cherry, 2000) in Florida and in countries with low disposable income, including Indonesia and Slovenia, where modest stakes by American standards represent several weeks' wages.

There are many interpretations of what causes Responders to reject substantial sums (see Chapter 3). There is little doubt that some players define a fair split of $10 as close to half and have a preference for being treated fairly. Such rejections are evidence of "negative reciprocity": Responders reciprocate unfair behavior by harming the person who treated them unfairly, at a substantial cost to themselves (provided the unfair Proposer is harmed more than they are). Negative reciprocity is evident in other social domains, even when monetary stakes are high—jilted boyfriends who accost their exes, ugly divorces that cost people large sums, impulsive street crimes caused by a stranger allegedly "disrespecting" an assailant, the failure of parties in le-

gal "nuisance cases" to renegotiate after a court judgment even when both could benefit (Farnsworth, 1999), and so on.[9]

This explanation for ultimatum rejections begs the question of where fairness preferences came from. A popular line of argument is that human experience in our ancestral past created evolutionary adaptations in brain mechanisms, or in the interaction of cognitive and emotional systems, which cause people to get angry when they are pushed around because getting angry had survival value when people interacted with the same people in a small group (see Frank, 1988). A different line of argument is that cultures create different standards of fairness, perhaps owing to the closeness of kin relations or the degree of anonymous market exchange with strangers (compared with sharing among relatives), and these cultural standards are transmitted socially through oral traditions and socialization of children.

Remarkable evidence for the cultural standards view comes from a study by eleven anthropologists who conducted ultimatum games in primitive cultures in Africa, the Amazon, Papua New Guinea, Indonesia, and Mongolia (see Chapter 2). In some of these cultures, people did not think that sharing fairly was necessary. Proposers in these cultures offered very little (the equivalent of $1.50 out of $10) and Responders accepted virtually every offer. Ironically, these simple societies are the *only* known populations who behave exactly as game theory predicts!

Note that rejections in ultimatum games do not necessarily reject the strategic principles underlying game theory (for example, Weibull, 2000). The Responder simply decides whether she wants both players to get nothing, or wants to get a small share when the Proposer gets much more. The fact that a Responder rejects means she is not maximizing her own earnings, but it does not mean she is not capable of strategic thinking. Recent theories attempt to explain rejections using social preference functions which balance a person's desire to have more money with their desire to reciprocate those who have treated them fairly or unfairly, or to achieve equality. Such functions have a long pedigree (traceable at least to Edgeworth in the 1890s). Economists have resisted them because it seems to be too easy to introduce a new factor in the utility function for each game. But the new theories strive to explain results in different games with a *single* function. Having a lot of data from different games to work with makes this enterprise possible and imposes discipline.

[9] My sister Jeannine told me that in Atlantic City the casinos sometimes have problems with lucrative "high-roller" customers stealing luxurious towels, robes, and other items from their (complimentary) hotel rooms after losing at the casinos. In their minds these losers are simply taking things they have paid for.

The new theories make surprising new predictions. For example, when there are two or more Proposers, there is no way for any one of them single-handedly to earn more money *and* limit inequality. As a result some theories predict that both Proposers offer almost everything to the Responder even though they *do* care about equality. (If there had been *two* photographers on that damn boat, we would have gotten our picture for $1.)

New social preference theories should prove useful in analyzing bargaining, tax policy, the strong tendency of tenant farmers to share crop earnings equally with landowners (Young and Burke, 2001), and wage-setting (particularly the reluctance of firms to cut wages in hard times, which is puzzling to economists who assume changes in the price of labor will equalize supply and demand, and other phenomena).

1.2.2 Example 2: Path-Dependent Coordination in "Continental Divide" Games

In coordination games, players want to conform to what others do (although they may have different ideas about which conformist convention is best). For example, in California there is an ongoing struggle over the physical location of the "new media" firms, such as internet provision of film and entertainment. New media people could gravitate toward Silicon Valley, where web geeks congregate, or toward Hollywood and Southern California, where many movies and TV shows are produced. Which geographical region is the better location depends on whether you think the location of internet firms is central, and "content" producers should follow them, or whether the internet is merely a distribution channel and content providers are king.[10]

This economic tug-of-war can be modeled by a game in which players choose a location, and their earnings depend on the location they choose and the location most other people choose. A game with this flavor has been studied by Van Huyck, Battalio, and Cook (1997). Table 1.1 shows the payoffs (in cents). In this game, players pick numbers from 1 to 14 (think of the numbers as corresponding to physical locations—low numbers are Hollywood and high numbers are Silicon Valley). The matrix in Table 1.1 shows the row player's payoff from choosing a number when the *median* number everyone in a group picks—the middle number—is the number in the different columns. If you choose 4, for example, and the median is 5, you earn a healthy payoff of 71; but if the median is 12 you earn −14 (bankruptcy!). The basic payoff structure implies you should pick a

[10] Of course, this example is undermined by the fact that cyberspace is everywhere and nowhere, so content providers might be able to stay put in the swank Hollywood Hills and still do business "in" Silicon Valley without moving.

low number if you think most others will pick low numbers, and pick a high number if you think most others will pick high numbers. If you aren't sure what others will do, pick a number such as 6, which gives payoffs ranging from 23 to 82 (hedging your bet).

In the experiments, players are organized into seven-person groups. The groups play together fifteen times. After each trial you learn what the median was, compute your earnings from that trial (depending on your own choice and the median), and play again. Since the game is complicated, think for a minute about what you would actually do and what might happen over the course of playing fifteen times.

The payoffs have the property that, if a player guesses that the median number is slightly below 7, her best response to that guess is to choose a number smaller than the guess itself. For example, if you think the median will be 7, your best response is 5, which earns 83 cents. Thus, if medians are initially low, responding to low medians will drive numbers lower until they reach 3. Three is an equilibrium or mutual best-response point because, if everyone chooses 3, the median will be 3 and your best response to a median of 3 is to choose 3. If players were to reach this point, nobody could profit by moving away. (The payoff from this equilibrium is shown in italics in Table 1.1.)

Table 1.1. *Payoffs in "continental divide" experiment (cents)*

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

Source: Van Huyck, Battalio, and Cook (1997).

But there is another Nash equilibrium. If players guess that the median will be 8 or above, they should choose numbers that are *higher* than their guesses, until they reach 12; 12 is also a Nash equilibrium because choosing 12 gives the highest payoff if the median is 12.

This is a coordination game because there are *two* Nash equilibria in which everybody chooses the same strategy. Game theorists have struggled for many decades to figure out which of many equilibria will result if there are more than one.

This particular game illustrates processes in nature and social systems in which small historical accidents have a big long-run impact. A famous example is what chaos theorists call the "Lorenz effect": Because weather is a complex dynamic system, the movement of a butterfly in China can set in motion a complicated meteorological process that creates a storm in Bolivia. If that butterfly had just sat still, the Bolivians would be dry! Another example is what social theorists call the "broken window effect." Anecdotal evidence suggests that, when there is a single broken window in a community, neighbors feel less obligation to keep their yards clean, replace their own broken windows, and put fresh paint on their houses. Since criminals want to commit crimes in communities where neighbors aren't watchful and other criminals are lurking (so the cops are busy), a single broken window can lead to a spiralling process of social breakdown. Policymakers love the broken window theory because it suggests an easy fix to problems of urban decay—repair every window before the effect of a few broken ones spreads throughout the community like a virus.

I call the game in Table 1.1 the "continental divide" game. The continental divide is a geographic line which divides those parts of North America in which water will flow in one direction from the parts in which water flows in the opposite direction. If you stand on the continental divide in Alaska, and pour water from a canteen as I once did, some drops will flow north to the Arctic Ocean and others will flow to the Pacific Ocean. Two drops of water that start out infinitesimally close together in the canteen end up a thousand miles apart.

The game is called the continental divide game because medians below 7 are a "basin of attraction" (in evolutionary game theory terms) for convergence toward the equilibrium at 3. Medians above 8 are a basin of attraction for convergence toward 12. The "separatrix" between 7 and 8 divides the game into regions where players will "flow" toward 3 and players will flow toward 12.

Which equilibrium is reached has important economic consequences. The 12 equilibrium pays $1.12 for each player but the 3 equilibrium pays only $0.60. On this basis alone, you might guess that players would choose higher numbers in the hopes of reaching the more profitable equilibrium. Before glancing ahead, ask yourself again what you think will happen. If you

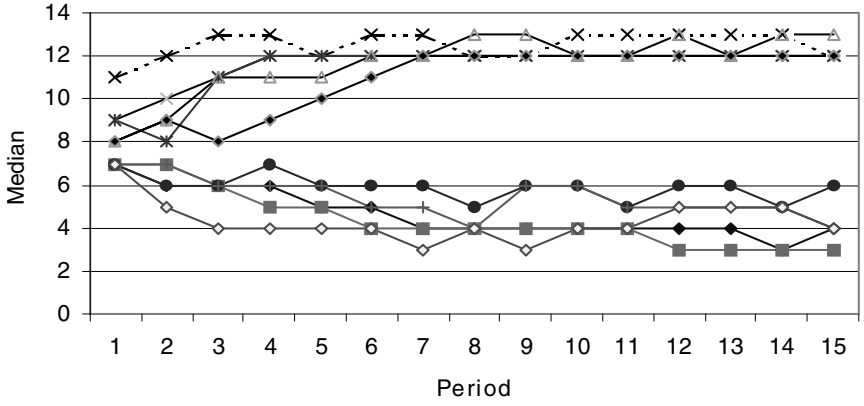


Figure 1.2. *Median choices in the "continental divide" game. Source: Based on data from Van Huyck, Battalio, and Cook (1997).*

have studied a lot of game theory and still aren't sure what to expect, your curiosity about what people actually do should be piqued.

Figure 1.2 shows what happened in ten experimental groups. Five groups started at a median at 7 or below; all of them flowed toward the low-payoff equilibrium at 3. The other five groups started at 8 or above and flowed to the high-payoff equilibrium.

The experiment has two important findings. First, people do *not* always gravitate toward the high-payoff equilibrium even though players who end up at low numbers earn half as much. (Whether they would if they could play again, or discussed the game in advance, is an interesting open question.) Second, the currents of history are strong, creating "extreme sensitivity to initial conditions." Players who find themselves in a group with two or three others who think 7 is their lucky number, and choose it in the first period, end up sucked into a whirlpool leading to measly $0.60 earnings. Players in a group whose median is 8 or higher end up earning almost twice as much. One or two Chinese subjects choosing 8—a lucky number for Chinese—could bring good fortune to everyone, just as the butterfly brought rain on the Bolivians.

No concept in analytical game theory gracefully accounts for the fact that some groups flow to 3 and earn less, while others flow to 12 and earn more. Indeed, the problem of predicting which of many equilibria will result in games such as these may be inherently unsolvable by pure reasoning. Social conventions, communication, subtle features of the display of the game, analogies players draw with experiences they have had, and homespun ideas about lucky numbers could all influence which equilibrium is reached. As

Schelling (1960) wrote, predicting what players will do in these games by pure theory is like trying to prove that a joke is funny without telling it.

1.2.3 Example 3: "Beauty Contests" and Iterated Dominance

In Keynes's famous book *General Theory of Employment, Interest, and Money*, he draws an analogy between the stock market and a newspaper contest in which people guess what faces others will guess are most beautiful: "It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree, where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth, and higher degrees" (1936, p. 156). This quote is perhaps no more apt than in the year 2001 (when I first wrote this), just after prices of American internet stocks soared to unbelievable heights in the largest speculative bubble in history. (At one point, the market valuation of the e-tailer bookseller Amazon, which had never reported a profit, was worth more than all other American booksellers combined.)

A simple game that captures the reasoning Keynes had in mind is called the "beauty contest" game (see Nagel, 1995, and Ho, Camerer, and Weigelt, 1998). In a typical beauty contest game, each of N players simultaneously chooses a number x_i in the interval [0,100]. Take an average of the numbers and multiply by a multiple $p < 1$ (say $p = 0.7$). The player whose number is closest to this target (70 percent of the average) wins a fixed prize. Before proceeding, think about what number you would pick.

The beauty contest game can be used to distinguish whether people "practise the fourth, fifth, and higher degrees" of reasoning as Keynes wondered. Here's how. Most players start by thinking, "Suppose the average is 50". Then you should choose 35, to be closest to the target of 70 percent of the average and win. But if you think all players will think this way the average will be 35, so a shrewd player such as yourself (thinking one step ahead) should choose 70 percent of 35, around 25. But if you think all players think that way you should choose 70 percent of 25, or 18.

In analytical game theory, players do not stop this iterated reasoning until they reach a best-response point. But, since all players want to choose 70 percent of the average, if they all choose the same number it must be zero. (That is, if you solve the equation $x^* = 0.7x^*$, you've found the unique Nash equilibrium.)

The beauty contest game provides a rough measure of the number of steps of strategic thinking that subjects are doing. It is called a "dominance-solvable game" because it can be "solved"—i.e., an equilibrium can be

computed—by iterated application of dominance. A dominated strategy is one that yields a lower payoff than another (dominant) strategy, regardless of what other players do. Choosing a number above 70 is a dominated strategy because the highest possible value of the target number is 70, so you can always do better by choosing a number lower than 70. But if nobody violates dominance by choosing above 70, then the highest the target can be is 70 percent of 70, or 49, so choosing 49–70 is dominated if you think others obey one step of dominance. Deleting dominated strategies iteratively leads you to zero.

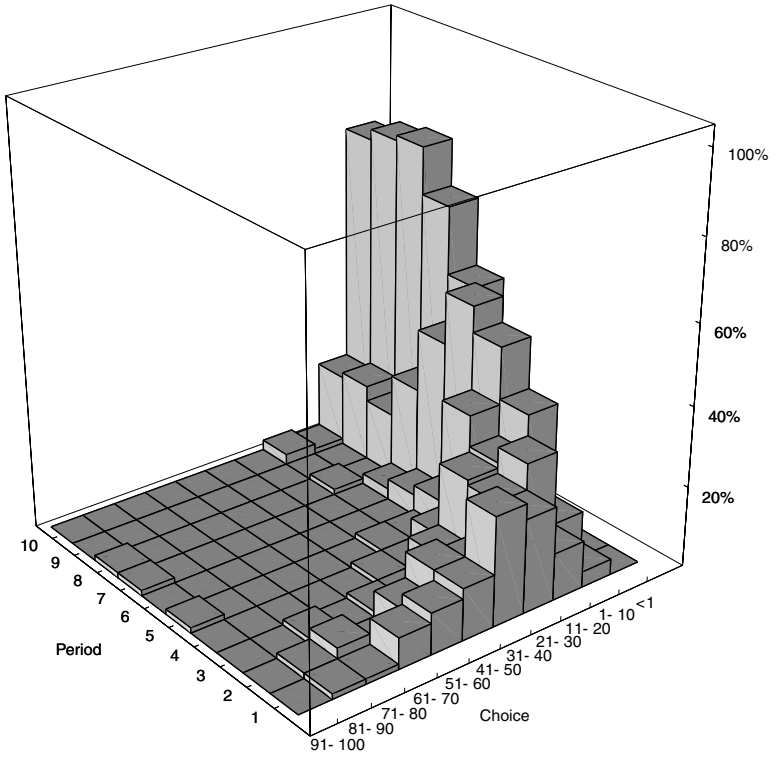
Many interesting games are dominance solvable. A familiar example in economics is Cournot duopoly. Two firms each choose quantities of similar products to make. Since their products are the same, the market price is determined by the total quantity they make (and by consumer demand). It is easy to show that there are quantities so high that firms will lose money because flooding the market with so much supply will drive prices too low to cover fixed costs. If you assume your rivals won't produce that much, then somewhat lower quantities are bad (dominated) choices for you. Applying this logic iteratively leads to a precise solution.

In practice, it is unlikely that people perform more than a couple of steps of iterated thinking because it strains the limits of working memory (i.e., the amount of information people can keep active in their mind at one time). Consider embedded sentences such as "Kevin's dog bit David's mailman whose sister's boyfriend gave the dog to him." Who's the "him" referred to at the end of the sentence? By the time you get to the end, many people have forgotten who owned the dog because working memory has only so much space.[11] Embedded sentences are difficult to understand. Dominance-solvable games are similar in mental complexity.

Iterated reasoning also requires you to believe that others are thinking hard, and are thinking that *you* are thinking hard. When I played this game at a Caltech board of trustees meeting, a very clever board member (a well-known Ph.D. in finance) chose 18.1. Later he explained his choice: He knew the Nash equilibrium was 0, but figured the average Caltech board member was clever enough to do two steps of reasoning and pick 25. Then why not pick 17.5 (which is 70 percent of 25)? He added 0.6 so he wouldn't tie with people who picked 17.5 or 18, and because he guessed that a few people would pick high numbers, which would push the average up. Now that's good behavioral game theory! (He didn't win, but was close.)

What happens in beauty contest games? Figure 1.3 shows choices in beauty contests with $p = 0.7$ with feedback about the average given to

[11] Seeing the sentence on the written page makes it easier; try reading it aloud to somebody who must remember the words and cannot refer back to them.

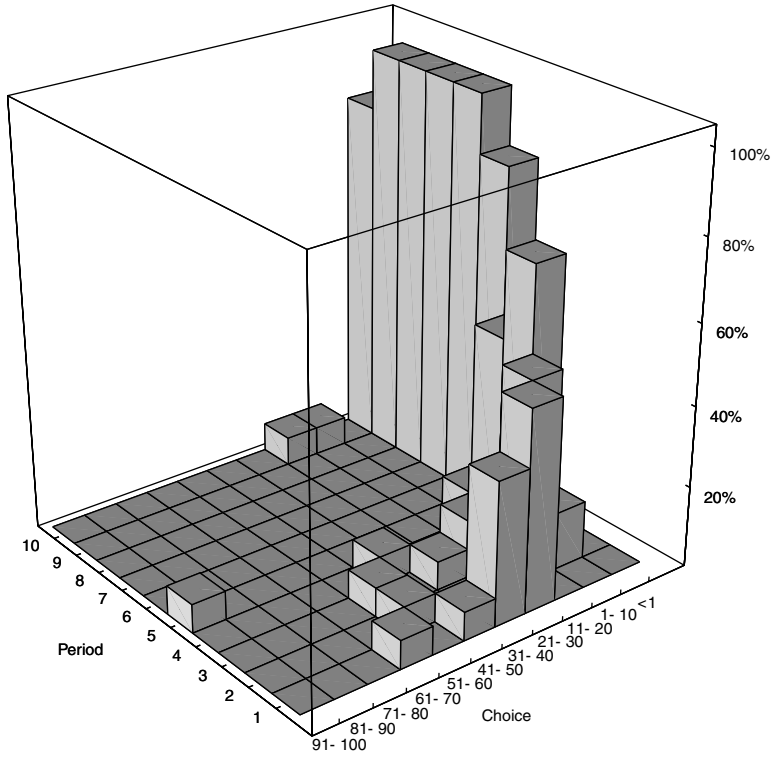


(a)

Figure 1.3. *Convergence in low-stakes and high-stakes "beauty contest" games. Source: Unpublished data from Ho, Camerer, and Weigelt.*

subjects after each of ten rounds (unpublished data from Ho, Camerer, and Weigelt). Bars show the relative frequency of choices in different number intervals (on the side) across ten rounds (in front). The first histogram shows results from games with low-stakes payoffs (a $7 prize per period for seven-person groups) and the second histogram shows results from high-stakes ($28) payoffs.

First-round choices are around 21–40. A careful statistical analysis indicated that the median subject uses one or two steps of iterated dominance. That is, most subjects roughly guess that the average will be 50 and choose 35, or guess that others will choose 35 and choose 25. Very few subjects chose the equilibrium of zero in the first round. In fact, they should *not* choose zero. The goal is to be *one* step ahead of the average but no further and choosing zero is being too smart for your own good!



(b)

Figure 1.3 (continued)

Although the game-theoretic equilibrium of zero is a poor guess about initial choices, players *are* inexorably drawn toward zero as they learn. Behavioral game theory uses a concept of limited iterated reasoning to understand initial choices and a theory of learning to explain movement across rounds.

The beauty contest has been replicated in dozens of subject pools (see Chapter 5 for details), including Caltech undergraduates,[12] trustees on

[12] Caltech students are a useful subject pool because they are extraordinarily analytically skilled. In many years, the incoming first-year class has a median math SAT score of 800. Recently, the average test scores of the *applicants* have been higher than the average of those students who are *accepted* at Harvard. Studying how these students play simple games establishes whether very analytical students can figure the games out. Generally they do not play much differently than students at other colleges.

the Caltech board (including a subsample of corporate CEOs), economics Ph.D.s and game theorists, and readers of business newspapers (the *Financial Times* in the United Kingdom, *Spektrum* in Germany, and *Expansion* in Spain). The results in all these groups are very similar: Players use 0–3 levels of reasoning, and few subjects choose the Nash equilibrium of zero. Comparing Figures 1.3(a) and 1.3(b) shows that increasing the prize by a factor of four, leading to average earnings of $40 for a 45-minute experiment, has only a small effect. (In the high-stakes condition there are more low-number choices in periods 5–10).

The limited iterated reasoning measured in these games provides one explanation for persistence of phenomena such as the stock price bubbles Keynes had in mind. Even if all investors foresee a crash, they do not "backward induct" all the way to the present. They guess that others will sell a couple of steps before the crash, and plan to sell just before that exodus. This reasoning process does not unravel all the way (because doubt "reverberates"), which explains why bubbles can persist even if everyone knows they will eventually burst. Allen, Morris, and Shin (2002) make their argument precise and Camerer and Weigelt (1993) and Porter and Smith, (1994) show that bubbles can happen in the lab.

1.3 Experimental Regularity and Behavioral Game Theory

This book is a long answer to a question game theory students often ask: "This theory is interesting . . . but do people actually play this way?" The answer, not surprisingly, is mixed. There are no interesting games in which subjects reach a predicted equilibrium immediately. And there are no games so complicated that subjects do not converge in the direction of equilibrium (perhaps quite close to it) with enough experience in the lab.

Consider the three examples above. In ultimatum bargaining, players are far from the perfect equilibrium-assuming self-interest, but they are roughly in equilibrium when the Responder's preference for being treated fairly is taken into account (because offers maximize expected profit given observed rejection rates). Behavioral game theory explains these results by combining new theories of social utility with analytical game theory (see Chapter 2). In the continental divide and beauty contest games, players start far from equilibrium and converge close to it in ten periods or so. Behavioral game theory explains these results using concepts of limited reasoning as players first think about a game (see Chapter 5) and precise theories of learning (see Chapter 6).

Sherlock Holmes said, "Data, data! I cannot make bricks without clay." Experimental results are clay for behavioral game theory. The goal is not to "*dis*prove" game theory (a common reaction of psychologists and sociolo-

gists) but to *im*prove it by establishing regularity, which inspires new theory. Without some sort of observation, theoretical assumptions are grounded in casual pseudo-empirical work—informal opinion polls in seminar and office discussions and using one's own intuitions (a one-respondent poll). Biologists don't just ask "If I was a robin foraging for food, how might I do it?" They watch robins forage, or ask somebody who has. Theorist (and part-time experimenter) Eric Van Damme, among others, worries about the effects of having too few data of this sort in game theory (1999, p. 204):

> Without having a broad set of facts on which to theorize, there is a certain danger of spending too much time on models that are mathematically elegant, yet have little connection to actual behaviour. At present our empirical knowledge is inadequate [precisely the same word von Neumann and Morgenstern used fifty years before!] and it is an interesting question why game theorists have not turned more frequently to psychologists for information about the learning and information processing processes used by humans.

Data are particularly important for game theory because there is often more than one equilibrium (see Chapter 7) and how equilibration occurs is not perfectly understood (see Chapter 6). Pure mathematics alone will not solve these problems.

Why has empirical observation played a small role in game theory until recently? One possibility is that early experimentation was thought to have "failed". In a 1952 RAND conference, several theorists (including eventual Nobel laureate Nash) gathered to think about game theory. They also did some experiments, the results of which did not confirm theory and reportedly discouraged Nash and perhaps others (Nasar, 1998).[13] Interest in data also suffered from the fact that so many interesting mathematical puzzles were open for solution in game theory for such a long time.[14] From about 1970 onward, developments in the theory of repeated games, games of incomplete information, and applications to important fields such as principal–agent relations, contracting, and political science led to an

[13] I think these early experimenters made a mistake by concentrating too much on games with mixed-strategy equilibria. In those games, players have low monetary incentives and predictions depend on assumptions about risk tastes, which are difficult to measure or even control.

[14] Many "modern" ideas in behavioral game theory were first proposed early in the history of game theory, and left aside or forgotten. In his thesis Nash (1950) described a "mass action" interpretation of equilibrium similar to modern evolutionary game theory (Weibull, 1995). Weighted fictitious play (see Chapter 6), which seems to have been revived by empiricists around 1995, is described in the amazingly insightful book by Luce and Raiffa (1957). Selten (1978) emphasized how players perceive the game they play, a topic being revived by Rubinstein (1991), Camerer (1999), and Samuelson (2001), among others. Rosenthal (1989) first proposed a "quantal response equilibrium" version, later refined and applied by McKelvey and Palfrey (1995, 1998) and Goeree and Holt (1999).

explosion of theory. There is no doubt that this pursuit has been extremely insightful and necessary, but it was conducted with little empirical guidance of any sort. There is also little doubt that it is high time to raise the ratio of observation to theory. It is also encouraging that some theorists have turned serious attention to modeling bounded or procedural rationality formally (e.g., Rubinstein, 1998).[15]

Of course, experimental data are only one component of behavioral game theory. Detailed facts about cognitive mechanisms and field tests are important too.[16] The result of controlled experiments, field observation, and theorizing working together is summarized by Vince Crawford (1997, p. 208):

> The experimental evidence suggests that none of the leading theoretical frameworks for analyzing games—traditional non-cooperative game theory, cooperative game theory, evolutionary game theory, and adaptive learning models—gives a fully reliable account of behavior by itself, but that most behavior can be understood in terms of a synthesis of ideas from those frameworks, combined with empirical knowledge in proportions that depend in predictable ways on the environment.

Rapid development of behavioral game theory will depend on how scientists react to data. Reactions vary.

If you are smitten by the elegance of analytical game theory you might take the data as simply showing whether subjects understood the game and were motivated. If the data confirm game theory, you might say, the subjects must have understood; if the data disconfirm, the subjects must have not understood. Resist this conclusion. The games are usually simple, and most experimenters carefully control for understanding by using a quiz to be sure subjects know how choices lead to payoffs. Furthermore, by inferring subject understanding from data, there is no way to falsify the theory. Physicists and biologists would not have the same reaction if a theory about particles were falsified by careful experimentation ("The particles were confused!") or if birds didn't forage for food as predicted ("If they had more at stake [than survival?] they would get it right!"). Game theorists should be similarly open-minded to what behaving humans can teach them about human behavior.

In fact, evidence cited as confirmation of game theory often supports a key element of *behavioral* game theory—namely, that equilibration may take a long time, perhaps years or decades (and equilibration is therefore a crucial component of any theory). In the foreword to Roth and Sotomayor's

[15] This includes finite automata, ϵ-equilibrium, evolutionary and dynamic theories, non-partitional information structures, and so on. Most of this work is not directly inspired or disciplined by data, however.

[16] Roth's work on matching for college bowl games, sorority rush, and medical residency are rare, impressive examples (e.g., Roth and Xing, 1994).

(1990) book about the theory of matching markets, the brilliant mathematician Robert Aumann notes that

> the Gale–Shapley [matching] algorithm had in fact been in practical use already since 1951 for the assignment of interns to hospitals in the United States; it had evolved by a trial-and-error process that spanned more than half a century. . . . in the *real* real world—when the chips are down, the payoff is not five dollars but a successful career, and people have time to understand the situation—the predictions of game theory fare quite well.

Note that the "time to understand the situation" Aumann refers to was fifty years![17] Over such a span, a learning or equilibration theory is essential.

Another reaction you may have is to criticize details of experimental design. Aumann, again, writes (1990, p. xi):

> It is sometimes asserted that game theory is not "descriptive" of the "real world," that people don't really behave according to game-theoretic prescriptions. To back up such assertions, some workers have conducted experiments using poorly motivated subjects, subjects who do not understand what they are about and are paid off by pittances; as if such experiments represented the real world.

Aumann is alluding to an earlier generation of experiments in the 1960s and 1970s which were not sensitive to subject comprehension and incentives. This book largely ignores those experiments (though some are described in Chapter 3). The modern experiments described in this book—mostly from the past ten years—fully respect concerns such as Aumann's and are designed with them in mind. Subjects are typically analytically skilled college students who are quizzed and highly motivated.

Another reaction you are likely to have when behavior does not conform to analytical game theory is that subjects were playing a different game than the experimenter created. Such explanations are useful if they can be tested and falsified. However, these explanations make experimenters bristle when they are made in ignorance of the extraordinary care taken to ensure subject comprehension, control for anonymity when trying to create one-shot games, and variation in stakes and subject pool to check for robustness.

[17] A similar point is made by Dixit and Skeath (1999). Stephen Jay Gould (1985) argued that baseball batting averages converged in the 20th century because of dynamic adjustments in field, pitching, and hitting. Dixit and Skeath describe this as an "encouraging tale, drawn from real life, of how players learn to play equilibrium strategies." But the learning was on the order of decades, which means a behavioral learning theory is just as important (or more so) than an equilibrium concept.

For example, a common interpretation of the fact that Responders reject offers in ultimatum games is that the Responders think they might be playing a repeated game because they will meet the Proposers again. But experimenters go to great lengths to ensure that subjects won't meet again and know that. For example, some experimenters pay subjects one at a time, with a short lag between each payment, and stand in the hall to be sure subjects don't wait for others to leave. Under these conditions, the faux-repeated-game explanation of ultimatum results is simply wrong. Others (such as the famously careful Ray Battalio) are known to end an experiment immediately if a subject says something aloud that others hear, breaking the experimenter's control. The reaction that subjects are playing a different game than the experimenter intended should disappear as more theorists learn about what actually happens in laboratories and come to believe in the quality of the data that are produced.

Still another reaction you may have is that behavior which is not rational can't be modeled. For example, several years ago Abreu and Matsushima (1992b) said experimental results are frequently inexplicable by "even approximately rational explanation." I disagree: Virtually all the results reported in this book can be accommodated by including behavioral components—social utility, limited iterated reasoning, and learning—into analytical theory. They go on to ask, "Should we then give up the rationality paradigm?" Of course not. It is too useful as a source of sharp predictions, and it is often a good prediction of limiting behavior. Behavioral game theory *extends* rationality rather than abandoning it. The last chapter of this book shows how.

1.4 Conclusion

This chapter described three examples which illustrate experimental regularity, and hinted how that regularity is formalized in behavioral game theory.

In the ultimatum game, Proposers typically offer close to half of a sum to be divided, and Responders reject offers that are too low because they dislike unfairness. The game is so simple that it is impossible to believe Responders rejecting money are confused, and the result has been replicated for very high stakes (up to $400 in America, and comparable sums in foreign countries). According to behavioral game theory, Responders reject low offers because they like to earn money but dislike unfair treatment (or like being treated equally). In the continental divide game, players gravitate toward equilibria over time and often end up in Pareto-inefficient equilibria they could have avoided. Behavioral game theory explains this by assuming that players aren't sure what to do (at the beginning of the game), so they

pick numbers in the middle; then they respond to history according to simple statistical learning rules. In the beauty contest game, players seem to do one or two steps of reasoning about others, then stop. (Analytical game theory assumes they keep going until they reach a mutual best-response equilibrium.) And they learn over time. Later chapters expand on these results and describe other classes of games (mixed equilibria, bargaining, signaling, and auctions).

APPENDIX

A1.1 Basic Game Theory

This appendix introduces basic ideas in game theory.[18] The goal is to equip the novice reader to understand the gist of the rest of the book. If you do not have some other background in game theory, and are serious about understanding the experimental results described later, you should read other books. A good introductory book (low on math) is Dixit and Skeath (1999). More mathematical books include Rasmusen (1994) and Osborne and Rubinstein (1995). Gintis (1999) includes fresh material on evolutionary theory and experimental data, and tons of problems. The heavy tomes that are used in graduate classes at places such as Caltech include Fudenberg and Tirole (1991).

Notation: Player i's strategy is denoted s_i. A vector of strategies, one for each player, is denoted $s = s_1, s_2, \ldots, s_n$. The part of this vector which removes player i's strategy (i.e., every other player's strategy) is denoted s_{-i}. The utility of player i's payoff from playing s_i is $u_i(s_i, s_{-i})$.

A1.1.1 Dominance

Definition A1.1.1 *The strategy s_i^* is a dominant strategy if it is a strict best response to any feasible strategy that the others might play*

$$u_i(s_i^*, s_{-i}) > u_i(s_i', s_{-i}) \quad \forall s_{-i}, s_i' \neq s_i^*.$$

The strategy s_i' is dominated if there exists $s_i'' \in S_i$ such that

$$u_i(s_i'', s_{-i}) > u_i(s_i', s_{-i}) \; \forall s_{-i}.$$

[18] Thanks to Angela Hung for writing much of this appendix.