# 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 knickknacks 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.)

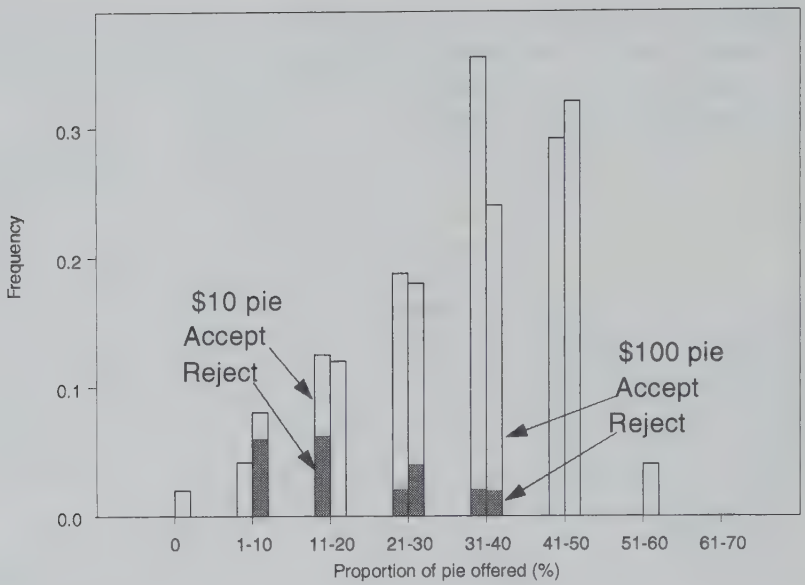



***Figure 1.1.*** *Offers and rejections in high- and low-stakes ultimatum games. Source: Based on data from Hoffman, McCabe, and Smith (1996a).*

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

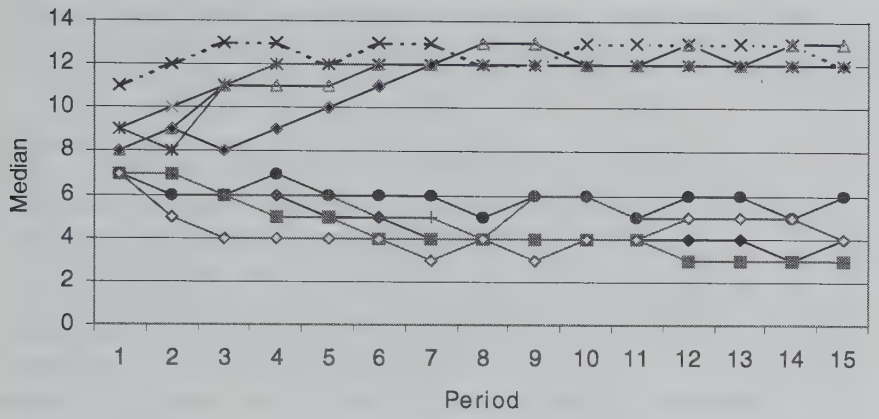



***Figure 1.2.*** *Median choices in the "continental divide" game. Source: Based on data from Van Huyck, Battalio, and Cook (1997).*

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.

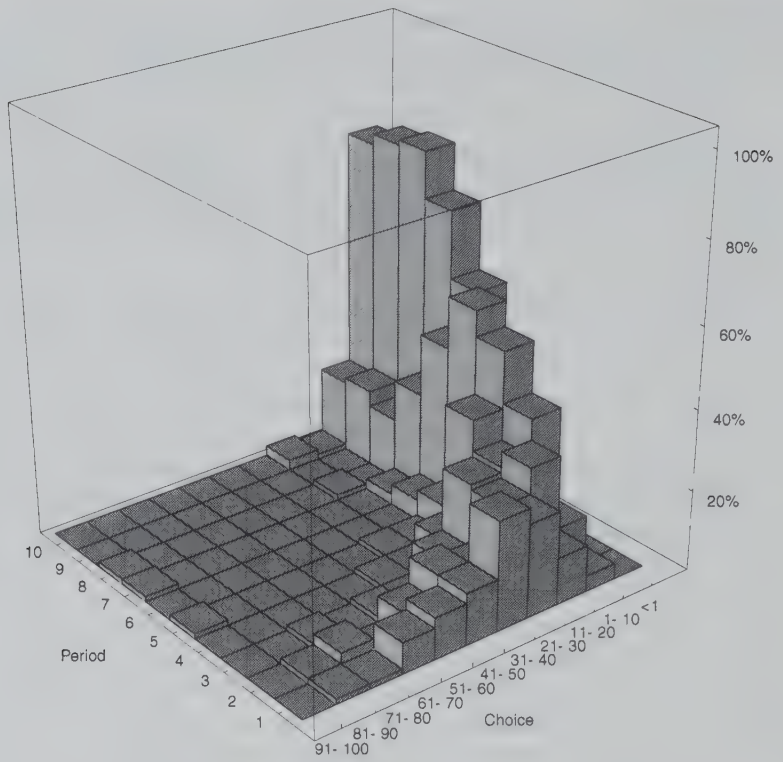
Many interesting games are dominance solvable. A familiar example in economics is Cournot duopoly. Two firms each choose quantities of similar products to make. Since their products are the same, the market price is determined by the total quantity they make (and by consumer demand). It is easy to show 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.

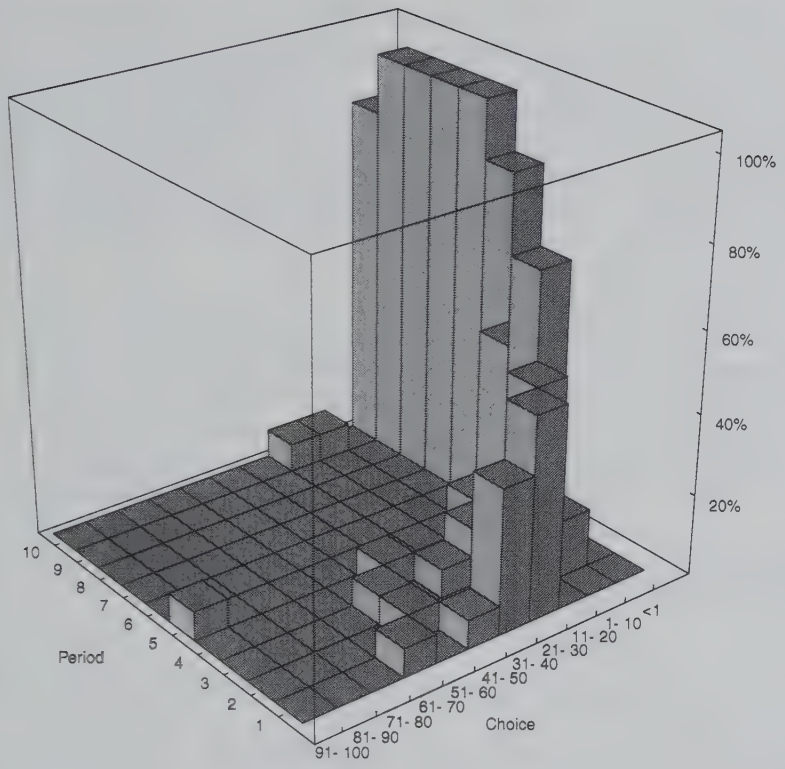



(a)

***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, $\epsilon$-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.

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

$$u_i(s_i'', s_{-i}) \geq u_i(s_i', s_{-i}) \ \forall s_{-i},$$
$$u_i(s_i'', s_{-i}) > u_i(s_i', s_{-i}) \text{ for at least one } s_{-i}.$$

***Example A1.1.1*** *Consider the simple normal-form game below. In a normal-form (aka strategic-form or matrix) game players are presumed to move simultaneously so there is no need to express the order of their moves in a graphical tree (or extensive-form). Each cell shows a pair of payoffs. The left payoff is for the row player (1) and the right is for the column player (2). The payoffs are* utilities *for consequences. That is, in the original game the consequences may be money, pride, reproduction by genes, territory in wars, company profits, pleasure, or pain. A key assumption is that players can express their satisfaction with these outcomes on a numerical utility scale. The scale must at least be ordinal—i.e., they would rather have an outcome with utility 2 than with utility 1—and when expected utility calculations are made the scale must be cardinal (i.e., getting 2 is as good as a coin flip between 3 and 1).*

| | | *Player 2* | | |
|---|---|---|---|---|
| | | *L* | *M* | *R* |
| *Player 1* | *U* | 1,0 | 1,2 | 0,1 |
| | *D* | 0,3 | 0,1 | 2,0 |

*For player 2, strategy R is strictly dominated by M (because M gives a higher payoff if player 1 chooses U, 2 instead of 1, and a higher payoff if player 2 chooses D, 1 instead of 0). Deleting strategy R (i.e., assuming a rational player 2 will never play it) makes D strictly dominated by U. But if player 1 plays U, then player 2 should play M. Therefore, the iterated-dominance equilibrium is (U,M).*

Dominance is important because, if utility payoffs are correctly specified (one need get only their *order* right) and players care only about their own utility, there is no good reason to violate strict dominance. One step of iterated dominance is a judgment by one player that the other player will not make a dumb mistake. This often tells a player what she herself should do. In the example, player 1 might consider choosing D because of the chance of earning the 2 payoff in the lower right (D,R) cell. But will she ever earn that payoff? Only if player 2 does something that is dominated. If player 1 assumes player 2 won't do that, she can rule out R, and her hope of earning 2 disappears. Then she should obviously choose U.

***Example A1.1.2*** *(Battle of the Sexes)*

| | | Player 2 | |
|---|---|---|---|
| | | *L* | *M* |
| *Player 1* | *U* | 2,1 | 0,0 |
| | *D* | 0,0 | 1,2 |

*The game is not dominance solvable. Neither strategy is dominant (or dominated) for either player because there is no one strategy that is always best. Put differently, each strategy* might *be best depending on what you think the other person will do.*

### *A1.1.2 Nash Equilibrium*

***Definition A1.1.2*** *The strategy profile* $s^* = (s_i^*, s_{-i}^*)$ *is a Nash equilibrium (NE) if each player's strategy is a best response to the other players' strategies. That is, no player has incentive to deviate, if no other player will deviate. (If players find themselves in equilibrium, there is no reason to move away.)*

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

Note that, if a strategy profile is an iterated-(strict) dominance equilibrium, then it is a Nash equilibrium. This is not true of equilibria created by iterated application of *weak* dominance.

***Example A1.1.3*** *(Battle of the Sexes) Solving for pure-strategy Nash equilibrium:*

| | | Player 2 | |
|---|---|---|---|
| | | *L* | *R* |
| *Player 1* | *U* | 2,1 | 0,0 |
| | *D* | 0,0 | 1,2 |

*If player 1 plays U, 2's best response is L. If player 1 plays D, 2's best response is R. If player 2 plays L, 1's best response is U, and if 2 plays R, 1's best response is D. Therefore, U is a best response to L, and L is a best response to U. Likewise, D is a best response to R and R is a best response to D: Pure strategy NE are (U,L) and (D,R).*

### *A1.1.3 Mixed Strategies*

A mixed strategy for player $i$ is a probability distribution over all the strategies in $S_i$.

***Example A1.1.4*** *(Battle of the Sexes) Solving for mixed-strategy Nash equilbrium:*

| | | Player 2 | |
|---|---|---|---|
| | | *L* | *R* |
| *Player 1* | *U* | 2,1 | 0,0 |
| | *D* | 0,0 | 1,2 |

*Suppose player 1 plays U with probability p and D with probability 1−p and player 2 plays L with probability q and R with probability 1−q.*

*Then the expected value to 2 from playing L is*

$$1p + 0(1-p)$$

*and the expected value to 2 from playing R is*

$$0p + 2(1-p).$$

*Player 2 is indifferent iff*

$$1p + 0(1-p) = 0p + 2(1-p)$$

*or*

$$p = \frac{2}{3}.$$

*The expected value to 1 from playing U is*

$$2q + 0(1-q)$$

*and the expected value from playing D is*

$$0q + 1(1-q).$$

*Player 1 is indifferent iff*

$$2q + 0(1-q) = 0q + 1(1-q)$$

*or*

$$q = \frac{1}{3}.$$

*As a result, a pair of (weak) best responses constitutes a mixed-strategy equilibrium:* $\left((\frac{2}{3}U, \frac{1}{3}D), (\frac{1}{3}L, \frac{2}{3}R)\right)$.

Mixed-strategy equilibrium is a curious concept. Introducing mixed strategies makes the space of payoffs convex (i.e., for any two points in the space, all points in between are in the space too), which is necessary to guarantee existence of a Nash equilibrium (in finite games). Guaranteed existence is a beautiful thing and is part of what makes game theory productive: For *any* (finite) game you write down, you can be sure to find an equilibrium. This means that a policy analyst or scientist trying to predict what will happen will *always* have something concrete to say.

However, the behavioral interpretation of mixing strategies is dubious. By definition, a player desires to mix only when she is indifferent among pure strategies, which means she does not (strictly) desire to mix with particular probabilities; she just doesn't care what she does. Furthermore, one player's equilibrium mixture probabilities depend *only* on the *other* player's payoffs, which is odd. A modern interpretation of mixed-strategy equilibrium (called "purification") is that one player might appear to be mixing but is actually choosing a pure strategy conditional on some hunch variable they privately observe. Mathematically, this works the same way—as long as each player's *belief* about the other players' choice matches the predicted probabilities, the mixed equilibrium is a mutual best-response point. Chapter 3 gives more detail.

***Example A1.1.5*** *A three-strategy example*

| | | *Player 2* | | |
|---|---|---|---|---|
| | | $L(r)$ | $M(s)$ | $R(1-r-s)$ |
| *Player 1* | $T(p)$ | 30,30 | 50,40 | 100,35 |
| | $M(q)$ | 40,50 | 45,45 | 10,60 |
| | $B(1-p-q)$ | 35,100 | 60,10 | 0,0 |

*Player 1:*

$$30r + 50s + 100(1-r-s) = 40r + 45s + 10(1-r-s)$$
$$= 35r + 60s + 0(1-r-s),$$

*or*

$$r = \frac{22}{83},$$
$$s = \frac{56}{83},$$
$$1 - r - s = \frac{5}{83}.$$

*Because the game is symmetric,*

$$p = r = \frac{22}{83},$$

$$q = s = \frac{56}{83},$$

$$1 - p - q = 1 - r - s = \frac{5}{83}.$$

### *A1.1.4 Constant-Sum Games*

In a constant-sum game, the sum of the payoffs of the players is constant across outcomes. Constant-sum games are actually extremely rare because even when the sum of physical payoffs is constant (like bargaining over money or food-sharing) the players' utilities probably do not add to a constant. For two-person, constant-sum games, minimax, maximin, and Nash equilibrium all select the same strategy.

***Definition A1.1.3*** *The strategy $s_i^*$ is a maximin strategy if it maximizes i's minimum possible payoff; that is,*

$$s_i^* = \arg\max_{s_i} \left[ \min_{s_{-i}} u_i \right].$$

***Definition A1.1.4*** *The strategy $s_i^*$ is a minimax strategy if it minimizes the other players' maximum possible payoff; that is,*

$$s_i^* = \arg\min_{s_i} \left[ \max_{s_{-i}} u_{-i} \right].$$

***Example A1.1.6*** *(Matching Pennies)*

| | | Player 2 | |
|---|---|---|---|
| | | *L* | *R* |
| *Player 1* | *U* | 1,0 | 0,1 |
| | *D* | 0,1 | 1,0 |

*If the game is expanded to include mixed strategies, then the maximin strategy for the row player 1 is to randomize equally over U and D, which gives an expected utility of 0.5 for both L and R. Hence, 0.5 is the maximin value. It is also the minimax strategy because it guarantees that the column player 2 makes no more than 0.5 (in expected utility). It is also the unique Nash equilibrium.*

In constant-sum games, minimax is a heuristic way of respecting the fact that the other player's best responsiveness (as in Nash equilibrium) will necessarily give you the lowest payoff, because the players' interests are strictly opposed. If you best-respond, I'll get the least; my best response to that likelihood is to maximize the least I can get.

### *A1.1.5 Extensive-Form Games and Information Sets*

An extensive-form game is used to model games where there is a specific order of moves. An extensive-form game is (1) a configuration of nodes and branches running without any closed loops from a single (root) starting node to its end (terminal) nodes; (2) an indication of which node belongs to each player; (3) probabilities that "nature" (an outside force) uses to choose branches at random nodes; (4) collections of nodes, which are called information sets; and (5) utility payoffs at each end node.

***Definition A1.1.5*** *An information set for a player is a collection of decision nodes satisfying:*

1. *The player has the move at every node in the information set, and*
2. *When the play of the game reaches a node in the information set, the player with the move does not know which node in the information set has been reached.*

### *A1.1.6 Subgame Perfection*

An equilibrium for an extensive-form game specifies what each player will do at each information set, even those not reached. Subgame perfection imposes the further restriction that players will actually play their equilibrium strategy if the subgame is reached (Selten, 1965). (A subgame is the continuation game from a singleton node—i.e., a node which has no other nodes in its information set—to the end nodes which follow from that node.) In my view, this "refinement" of Nash equilibrium simply patches up an omission in Nash's concept, which was not evident until theorists began thinking about extensive-form trees rather than matrices.

***Example A1.1.7*** *Mini-ultimatum game*

| | | *Player 2* | |
|---|---|---|---|
| *Player 1* | *E* | 5,5 | |
| | | *A* | *R* |
| | *U* | 8,2 | 0,0 |

In this mini-ultimatum game, player 1 moves first and offers an even (E) split of 10, paying (5,5), or offers an uneven (U) split. If E is chosen, the game ends and both players earn 5. If U is chosen, player 2 is "on the move" and can choose to accept (A), in which case player 1 gets 8 and player 2 gets 2, or can reject (R), in which case both get nothing.

The strategy profile (E, R|U) is a Nash equilibrium because it specifies moves at each node, and strategies are—technically—best responses. If player 1 anticipates that player 2 will choose R after a move of U, then player 1 should choose E to earn 5. And if player 1 is going to choose E, it makes no difference what player 2 does—"planning" to respond with R to U is not penalized because, in equilibrium, player 2 is never called upon actually to play. Playing U after R is a weak best response because in this equilibrium R never results. Subgame perfection requires that, if the U node *is* reached, then player 2's subsequent strategy must be a best response. But since R earns 0 for player 2 and A earns 2, the best response in the subgame that results when player 1 chooses U is to play A. Anticipating this, player 1 should choose U. Therefore, (U, A|U) is a Nash equilibrium and is also subgame perfect.

### *A1.1.7 Bayesian-Nash Equilibrium*

In games of incomplete information, at least one player is uncertain about the other players' payoff function(s). This is traditionally represented (after Harsanyi, 1967–68) by having "nature" move at the beginning of the game and determine a player's "type." Players observe their own types but not the types of others. (In formal terms, the player who knows her type knows which branch emanating from the start node nature chose. The player who does not know the other player's type has an information set containing nodes that emanate from both branches following from the start node.) However, the probability distribution of types is common knowledge (the "common prior" assumption). Bayesian-Nash equilibrium adds two features to Nash equilibrium: (i) Along the equilibrium path (i.e., for all moves that occur in equilibrium with positive probability), players must update their beliefs about player types using Bayes' rule. Bayes' rule states that $P(H_i|D) = P(D|H_i)P(H_i)/\sum_k^n = 1P(D|H_k)P(H_k)$, where $H_i$ are $n$ different hypotheses (e.g., possible player types) and $D$ is the observed data (e.g., a player's move). (ii) Off the equilibrium path (i.e., after moves that never occur in equilibrium), players should have *some* belief about player types. Note well that Bayes' rule does not restrict what these beliefs should be, because an off-path move has probability zero (i.e., $P(D|H_i) = 0\ \forall H_i$). Then the denominator of the updating equation is zero and Bayes' rule breaks down. Hence, Bayesian-Nash equilibrium imposes the minimal a-Bayesian restric-

tion on what off-path beliefs should be—namely, they should be *something*! This simply rules out the possibility that players will violate dominance after observing an off-path move.

### *A1.1.8 Trembling Hand Perfection*

Selten (1975) suggested a clever way to subject off-equilibrium-path beliefs to the discipline of Bayes' rule, called "trembling hand perfection." The idea is to suppose that, even in an equilibrium, there is a small chance that a player's hand trembles when she chooses, so that *all* paths through the tree are taken with positive probability. Then Bayes' rule can be used to update beliefs. A trembling hand perfect equilibrium is the limit of the Bayesian-Nash equilibria with trembling, as the tremble probability goes to zero. Many others have suggested further refinements which try to codify, logically, what sort of beliefs after off-path moves seem intuitive or sensible. Sequential equilibrium (Kreps and Wilson, 1982a) is a kissing cousin of trembling hand perfection and is generically the same (i.e., the only games in which the two differ are knife-edge cases, in a way that can be made mathematically precise). Myerson (1978) suggested that the tremble probabilities should be smaller when payoff differences between equilibrium and nonequilibrium strategies are larger, leading to a concept of "proper" equilibrium. Some other refinements are discussed in Chapter 8, on signaling games.

### *A1.1.9 Quantal Response Equilibrium*

In a quantal response equilibrium (QRE), players do not choose the best response with probability one (as in Nash equilibrium). Instead, they "better-respond," and choose responses with higher expected payoffs with higher probability. In practice, the QRE often uses a logit or exponentiated payoff response function:

$$P(s_i) = \exp\left(\lambda \sum_{s_{-i}} P(s_{-i}) u_i(s_i, s_{-i})\right) \Big/ \sum_{s_k} \exp\left(\lambda \sum_{s_{-i}} P(s_{-i}) u_i(s_k, s_{-i})\right),$$

where $\exp(x)$ denotes $e^x$ and the sums are taken over all strategies for $-i$ (all other players) and $i$. Intuitively, QRE says that players fix a strategy and form beliefs about what others will do ($P(s_{-i})$), and compute expected payoffs given those beliefs. Making this calculation for each strategy gives a profile of expected payoffs for each possible strategy. Then player $i$ better-responds by choosing noisily according to the strategies' expected payoffs. The parameter $\lambda$ is a measure of their sensitivity to differences in expected

payoffs.[19] Note that since player I is calculating $P(s_i)$, and others are too, the system of equations is recursive: A player's behavior determines expected payoffs, which determine other players' behavior. When $\lambda = 0$, players just choose each strategy with the same equal probability. As $\lambda$ rises, they become more and more responsive, converging to Nash equilibrium in which they always choose the best response.[20] Thus, Nash equilibrium is a kind of "hyperresponsive" QRE.

I used to say in classes and seminars that, if John Nash had been a statistician rather than a mathematician, he might have discovered QRE rather than Nash equilibrium. (Such an early discovery would have automatically presolved the problem of refining away incredible Nash equilibria, which required the development, much later in the 1960s and 1970s, of subgame and trembling-hand perfection.) When I mentioned this at a talk in Princeton in the fall of 2001, some audience members grinned and nudged each other (Nash was, after all, a hometown hero in Princeton). People later said they were grinning and nudging because, unbeknownst to me, Nash had been in the audience! Since Nash didn't *protest* when he heard my counterfactual speculation about his would-be early discovery of QRE, I later stretched the truth and said "Nash didn't *deny* that he would have discovered QRE." Still later, in December 2001, I had a chance to meet Nash and asked him point blank about QRE. He said he had been working on a similar stochastic best-response model of bargaining just recently; so we can count him as recently converted to, or at least sympathetic toward, a quantal response approach.

## A1.2 Experimental Design

The way in which an experiment is conducted is unbelievably important. Just as all thoroughbred racehorses are descended from four horses, most American experimental economics began in the 1960s and 1970s at a small number of institutions (particularly Caltech, Arizona, Purdue, and Texas A&M) and grew slowly. (A similar effort occurred in parallel in Germany.) As a result, the experimental community is tight-knit and has established clear conventions for experimental practice which permit a high degree of comparability across data sets. Smith (1976) is an early rulebook which also summarizes many regularities. For example, most articles include raw data and instructions to enable readers to judge for themselves what was learned. (If you asked a psychologist for data or instructions he or she might

[19] Goeree and Holt (1999) use $1/\mu$ instead to emphasize their interpretation of the noise as computational mistakes; when $\mu$ is large, the mistake rate is high and players are very insensitive, and vice versa.

[20] This is not quite true, technically. The limit of a sequence of QREs as $\lambda$ increases can converge to something that is not a Nash equilibrium (see McKelvey and Palfrey, 1995).

be insulted, because the convention in that field is to give the writer the benefit of the doubt.)

This appendix sketches some important design choices. To learn more, see Friedman and Sunder (1993), Davis and Holt (1993), and Kagel and Roth (1995).

### *A1.2.1 Control, Measure, or Assume*

Any variable can be evaluated in one of three ways: control, measurement, or assumption.

Control means taking an action to affect the value of a variable, often with a "manipulation check" to be sure the control worked. Induced value is an important kind of control which creates preferences for actions by associating those actions with payoffs in a currency that subjects value (typically money, but sometimes grade points, ranking of points earned, and so forth).

Measurement refers simply to measuring the value of a relevant variable through psychometric measures ("Describe how angry you are when someone offers you $2 out of $10?"), methods for measuring risk-aversion (e.g., certainty equivalents) or probability judgments (scoring rules). Types of measurement that are less familiar in economics, but worth exploring, include content analysis of videotapes, physiological measures such as heartrate or galvanic skin response, information acquisition (see Johnson et al., 2002, in Chapter 4, and Costa-Gomes, Crawford, and Broseta, 2001, in Chapter 5), and even fMRI brain imaging (Smith et al., 2002).

Assumption is pseudo-control in which the experimenter is willing to accept a maintained hypothesis about the value of a variable.

As an illustration of all three methods, consider the classic economic experiments in which agents are endowed with costs and valuations for an object, and one would like to test theories of competitive equilibrium (CE), which predict that prices will converge to the point where supply meets demand. One strategy is to hand out everyday objects, such as CD recordings or coffee mugs, and make an assumption about how much subjects value them. This is generally a bad design because CE predictions are very sensitive to the valuations of marginal traders, and the valuations are not likely to be well understood by experimenters.[21] Another strategy is to

[21] A beautiful exception, which proves the rule, is the Kahneman, Knetsch, and Thaler (1990) experiments with coffee mugs. They were interested not in prices at all, but only in the quantity of trade, and in showing that aversion to losses creates an "endowment effect" that is present with everyday objects and *not* with induced-value tokens. They were able to use everyday objects because they did not care about the level of homemade valuations, but needed only to assume that mean valuations were similar in samples that (randomly) did and did not receive mugs.

measure valuations for each individual (using, say, the incentive-compatible Becker–DeGroot–Marschak procedure) and then assume those measured valuations represent the subjects' costs and reservation prices and construct demand and supply curves from the measurements. This is a defensible procedure but has rarely been used (see Knez and Smith, 1987). A third strategy is to "induce" or control valuations by making the objects of trade valueless tokens, and telling subjects that they can trade tokens for specific money values. This form of induced valuation, first used by Chamberlin (1948) and tirelessly refined by Vernon Smith beginning in 1956, is surely the crux move in the development of experimental economics. Smith's later insistence on actually tying money payments to the induced valuations led to credibility among nonexperimenters and reliability in actual behavior, which enabled exploration of extremely subtle hypotheses and rapid progress.

### *A1.2.2 Instructions*

Instructions tell subjects what they need to know. It is scientifically very useful to have a clear instructional "script" that enables precise replication, particularly across subject pools who may vary in language comprehension, obedience, intrinsic motivation, and so on. (Precise replication is surprisingly rare in other social sciences, in psychology for example.) Reading instructions out loud is a common practice to establish "public knowledge" (e.g., what subjects know everyone else has been told, and what everyone else knows everyone else has been told), which is as close as we can practically come to the common knowledge usually assumed in game theory.

The overwhelming convention in modern (post-1975) game theory experiments is to explain how each sequence of moves by each player leads to payoffs (including payoffs to other subjects, and also including asymmetric information à la Harsanyi). This practice arose because experimenters wanted to be sure that subjects had enough information to compute an equilibrium. (Earlier experiments on markets deliberately withheld information about values of others, to test the Adam Smith/Hayek hypothesis that, even if players knew only their own values, they could still converge to a Pareto-efficient equilibrium.) More recently, learning models have been proposed that do not always assume players have complete knowledge of all possible payoffs. As an empirical matter, it is quite interesting to know how people learn in these environments (since players may have poor information about payoffs in many naturally occurring situations). So some recent experiments have deliberately withheld information about payoffs from the instructions (e.g., Van Huyck, Battalio, and Rankin, 2001). This design choice also raises

a problem—when subjects are *not* told something about the environment they are placed in, their default assumption may be wrong.

Here is an example. A large literature on "probability matching" studies subjects (including nonhuman animals) making one of two choices (left L and right R). On each trial one of the levers is "armed" to deliver a reward and the other delivers nothing. Subjects choose one lever and receive a reward if it is armed. Typically, there is a chance $p$ that L is armed and $1 - p$ that R is armed, and *which lever is armed is determined independently of previous trials*. From the experimenter's view, the rewards from choosing L are independent Bernoulli trials. But what do subjects think? If they are not told that the lever-arming process is independent and identically distributed (with fixed $p$), they might entertain an array of possibilities of how rewards are delivered. It might be much more plausible, to a subject, that the experimenter is interested in whether they can figure out an elaborate pattern of variation in which levers are armed rather than figure out that L has an independent $p$ chance each time. Empirically, subjects typically "probability match" by choosing L on about $p$ of the trials. If subjects knew the process was independent and identically distributed, this would be a mistake (they should always choose the arm with the higher $p$). Does probability matching tell you that subjects are making a mistake, or that they fail to guess the strange (uninteresting) environment the experimenter has placed them in? What the experimenter observes in a low-information probability-matching experiment is a combination of the subjects' revealed perceptions about the statistical process of reward and their decision rule for choosing. Withholding statistical information is not necessarily a bad design choice. But, by not revealing information about the environment, there is no guarantee that subjects will guess accurately what they are not told. Since the guesses are hard to observe directly (without additional measurement), it will be difficult to conclude whether subjects are playing rationally or not.

### *A1.2.3 Anonymity*

If the subjects know the identity of the person they are bargaining with, their knowledge might influence what they do for many reasons. They might like the way the person looks and want to make them happy, or fear retribution or embarrassment if they make a stingy offer and see the person after the experiment. Unless these possibilities are precisely the focus of the experiment, most of the experiments described in this book try to create anonymity—sometimes to a dramatic degree—by making it as difficult as possible for subjects to know precisely who they are playing with. Anonymity is obviously *not* used because it is lifelike. It is used to establish a benchmark

against which the effects of knowing who you are playing can be measured, if those effects are of interest.

### *A1.2.4 Matching Protocols and Reputation Building*

Experiments are usually designed with several periods of play so subjects can learn from experience. But if a pair of subjects play together several times, the possibility of "reputation building" can affect the prediction that game theory makes. For example, in the ultimatum game it may pay for a Responder to build up a reputation for being "tough" by rejecting large offers in the first few periods, to "teach the proposer a lesson" and get larger offers in later periods. Put more formally, when the same pair (or group) of players play together several times, there may be game-theoretic equilibria for the repeated game that differ from the one-shot, stage-game equilibrium. Unless we are explicitly interested in the nature of reputation formation and repeated-game strategies (as some experiments are), this possibility is avoided by having subjects play with each other only once in an experimental session.

There are various ways in which players can be matched with different players in a stage game that is repeated. The most common protocol is no-repeat rematching (or a "stranger" design)—players are never rematched with a former match, to reduce the possibility of reputation building. (It is not known, by the way, whether no-repeat rematching actually does disable reputation building. I suspect it does not, but more work on this is needed.) In a "no-contagion" design, players are never rematched, and are never rematched with somebody who will be matched with somebody they will later be matched with, and so forth. In random rematching, players are rematched randomly (so they may be rematched with their partner from the previous period, but typically the probability of consecutive matches is low and they do not know whether they are rematched anyway). In a mean-matching or population protocol, each player plays every other player and earns the average payoff from all those matches.

### *A1.2.5 Incentives*

Vernon Smith (1962) reported the earliest experiments comparing the behavior of subjects who were rewarded in points with that of subjects whose points were converted into dollars that they were actually paid. Vernon observed that subjects paid only in points tended to approach competitive equilibrium more erratically, and seemed to grow bored with the experiment faster than those who were paid money. He suggested that, although people may have enough intrinsic motivation to earn lots of points, hypothetical rewards were typically more "erratic, unreliable, and easily satiated"

(1976) than money. Put differently, by inducing value using money payments, the experimenter need rely only on the assumptions that everybody likes having more money and nobody gets tired of having more of it. These are safe assumptions, and substantially safer than figuring out whether somebody is motivated by having their name posted if they did best (some people might be embarassed by it), is likely to give up if they are far behind when payoffs have a tournament structure, and so forth. (If you know anybody who is tired of getting more money let me know; I'll take their leftovers!)

Paying subjects their earnings quickly became the norm in experimental economics (in sharp contrast to most of modern experimental psychology, with important exceptions such as Ward Edwards and Amnon Rapoport). Does it matter whether performance is rewarded, and how much? The evidence is mixed. Experimenters should not abandon the practice of paying performance-based incentives, but few results that disconfirm theory have been overturned by paying more money. Smith and Walker (1993) review a couple of dozen studies and argue that paying money reduces variance of responses around a rational prediction (first noted by Dave Grether in a 1981 working paper). Hogarth and I (Camerer and Hogarth, 1999) conducted a more thorough review and draw several conclusions. Paying money *does* reduce variation and outliers, which may be particularly important in settings that are sensitive to variation, such as "weak-link" coordination games (see Chapter 7) or asset markets with potential for speculative bubbles. (In those tasks investigators should certainly pay money.) Paying money improved performance most reliably in judgment and decision-making tasks, when there are returns to thinking harder (see also Hertwig and Ortmann, 2001). But in tasks that are quite easy ("floor effects") or very hard ("ceiling effects") paying money usually does not matter. We also point out that there is no empirical reason to obsess only about money, because the effect of experience is just as large. Labeling strategies, individual differences, and other variables can have comparably large effects and should be investigated further.

In the 1980s a controversy erupted over whether money payments established what Vernon Smith called "dominance," which means that the money at stake is enough to induce subjects to think hard. The controversy was ignited by Harrison (1989), who pointed out that the size of a deviation from theory in payoff terms may be much lower than the deviation in the strategy space. (The same point—the "flat maximum critique"—was made almost two decades earlier, in 1973, by Von Winterfeldt and Edwards.) For example, in a mixed-strategy equilibrium, if other players are using their mixture strategies, then a subject has absolutely no financial incentive to play her equilibrium mixture instead, regardless of the level of money payoffs in the game. Similarly, in first-price private-value auctions, a bidder who overbids by, say, $1 does not actually lose a dollar. Overbidding reduces her prospective earnings but raises the chance of winning the auction, and the net effect may reduce her expected payoff by only pennies.

Experimenters never developed an ideal solution to the flat maximum problem. The critique did sensitize us to the need to worry about marginal costs of deviation (and see Fudenberg and Levine, 1997) and to seek designs with steep marginal incentives where possible. (For ultimatum game Responders, for example, the cost of the error from rejecting is exactly equal to the deviation in strategy space, and is often very large.) Furthermore, the fact that very large variations in stakes typically have only modest effects muted criticisms that payoffs were not large enough. At least two dozen studies have been done in foreign countries where purchasing power is so low that modest sums by the standards of developed countries amount to several weeks' or months' earnings. The results are generally very close to those with smaller stakes.

Finally, Rob Kurzban mentioned a great example of poor reasoning in a very high-stakes situation. In the final round of the first edition of the popular *Survivor* television show, survivor Greg Buis was deciding which of two others to vote for based on their answers in a simple number game. (The winner got $1 million and the runner-up $100,000, so the stakes are huge.) The two finalists, Richard Hatch and Kelly Wigglesworth, were asked to pick integers between 1 and 9 and the person whose number was closest to Buis's would win. (Whether Buis actually committed to his number, or simply used the contest to create an illusion of fairness, is hard to tell.) Hatch chose 7. Astonishingly, Wigglesworth then chose *3*, a choice that is dominated by choosing 6, because 6 would win for any choice by Buis of 6 or below whereas 3 would lose if Buis picked 6 and tie if he picked 5. If Buis had picked randomly, by choosing 3 after Hatch's 7 Wigglesworth lost an expected $160,000. And, unless one had a reason to think Buis would go high or low, picking 5 would have been a better choice by Hatch than 7. (Buis claimed he chose 9 and so Hatch won the $1 million.)

### *A1.2.6 Order Effects*

Experiments often involve two treatments, A and B. If they are always done in the same order, denoted AB, then any difference in the two treatments might be due to the fact that A came first and B came second (an "order effect"). Order is "confounded with" (perfectly correlated with) the treatment. This is easily controlled by running some sessions in the reverse order, BA, and including an order dummy variable in statistical analyses.

### *A1.2.7 Controlling Risk Tastes*

Even though subjects should be risk-neutral toward small lab gambles (Rabin, 2000), it would be useful to have a procedure for creating payoffs that subjects are risk-neutral towards (i.e., so they are indifferent to the disper-

sion of possible payoffs around a fixed mean). There *is* such a procedure—the binary lottery procedure.

In the binary lottery procedure, subjects are paid in lottery tickets, which are later used to determine their chance of winning a lottery for a fixed prize (see Roth and Malouf, 1979). If players reduce compound lotteries to single-stage lotteries, they should be neutral toward mean-preserving spreads in their ticket distribution—that is, they should have linear utility over tickets. For example, if they regard a 0.32 chance of winning fifty tickets (and otherwise getting none) as the same as having sixteen tickets, they are risk-neutral toward tickets. (In theory, the procedure can be extended to induce any shape of utility function by transforming payoff units to tickets nonlinearly; see Schotter and Braunstein, 1981, and Berg et al., 1986.)

Unfortunately, there is little evidence that the binary lottery procedure works as it should in theory (and a couple of studies showing it does not work). For example, in direct tests, players who make choices with monetary payoffs and players who make choices with lottery ticket payoffs exhibit the same patterns, so the binary lottery procedure does not change apparent risk-aversion over money into risk-neutrality over tickets (Camerer and Ho, 1994; Selten, Sadrieh, and Abbink, 1999). On the other hand, Van Huyck and Battalio (1999) found that players behaved consistently with risk-neutrality over tickets (though see their footnote 9 for an opposite result). In the most careful study, Prasnikar (1999) found that risk-aversion coefficients estimated from choices among gambles over tickets were close to their predicted coefficient values. The method worked best for the minority of subjects who obeyed reduction of compound lotteries.

It is surprising that many experimenters use the binary lottery procedure despite so little careful evaluation of when it does induce risk-neutrality (and given the evidence that it often doesn't). To paraphrase G. B. Shaw's wisecrack about marriage, faith in the procedure seems to be a triumph of hope over data. There are two alternatives to trying to induce risk tastes: assume risk-neutrality, or measure risk tastes over money independently and use those measures to calibrate an individual subject's risk preferences in a game (which several experimenters have done). In any case, it would be good to see more careful research (à la Prasnikar, 1999) establishing when the procedure works and when it does not.

### *A1.2.8 Within-Subjects and Between-Subjects Design*

In a "within-subjects" design, a single subject is observed in different treatments. (The subject serves as "her own control group".) In a "between-subjects" design—the norm in experimental economics—different subjects are tested in treatments A and B. Statistical variation across the subjects

then muddies the waters of what is observed by comparing A and B. Within-subject designs are more statistically powerful than between-subject designs because they automatically control for individual differences, which are often a large source of variation, and hence allow the effect of a treatment to shine through when the nuisance of individual difference is controlled for.

There is a curious bias against within-subjects designs in experimental economics (not so in experimental psychology). I don't know why there is a bias, and I can't think of a compelling reason always to eschew such designs. One possible reason is that exposing subjects to multiple conditions heightens their sensitivity to the differences in conditions. This hypothesis can be tested, however, by comparing results from within- and between-subjects designs, which is rarely done.

### *A1.2.9 Experimetrics*

"Experimetrics" are econometric techniques customized to experimental applications. Although I'm an amateur econometrician, I am a huge fan of experimetrics. The next generation of experimenters should feel obliged to use the very latest inferential tools—the best microscopes—to see patterns in data as clearly as possible. The work by Crawford, El-Gamal, McKelvey and Palfrey, Stahl, and Van Huyck and Battalio described in this book sets a high standard other experimenters should emulate.

A new tool in experimetrics which has not become widely adopted is optimal endogeneous experimental design (e.g., El-Gamal, McKelvey, and Palfrey, 1993). In many experiments, the experimenter has one or more hypotheses that she can put prior probabilities on. A crisp prior, and specification of the hypotheses, can be used to compute the information value (in the sense of dispersion of posterior probabilities relative to dispersion of priors) of different experimental design parameters. This motivates the choice of "optimally informative" design parameters. Furthermore, because of increases in computing power, for the first time in human history we can alter the experimental design in real time—while subjects are waiting, for seconds rather than days—to optimize the amount of information collected in an experiment. (Seen this way, all previous experimental designs are heuristic approximations to endogeneously optimized designs.) Information-optimized designs have rarely been used. The younger generation should embarass us older folks by taking them up with a vengeance.

# 2
# Dictator, Ultimatum, and Trust Games

IN 1982, GÜTH, SCHMITTBERGER, AND SCHWARZE reported the kind of empirical finding that surprises only economists. They studied an "ultimatum" game in which one player, the "Proposer," makes a take-it-or-leave-it offer, dividing some amount of money between herself and another person. If the second person, the "Responder," accepts the division, then both people earn the specified amounts. If the Responder rejects it, they both get nothing. The ultimatum game could hardly be simpler. If Responders maximize their own money payoffs, they should accept any offer. If Proposers also maximize and expect Responders to maximize, they should offer the smallest amount.

In experiments, Proposers offer an average of 40 percent of the money (many offer half) and Responders reject small offers of 20 percent or so half the time. The data falsify the assumption that players maximize their own payoffs as clearly as experimental data can. Every methodological explanation you can think of (such as low stakes) has been carefully tested and cannot fully explain the results.

Since the equilibria are so simple to compute (the Responder's move is just a choice of a payoff allocation), the ultimatum game is a crisp way to measure social preferences rather than a deep test of strategic thinking (see Marwell and Schmitt, 1968). Measuring social preferences in money terms is important because concepts such as fairness and trust figure prominently in private negotiation and public policy. But many cynics (especially economists) think fairness is simply a rhetorical term used by people who deserve the short end of the stick for trying to get more, and that people will not sacrifice much to punish unfairness or reward fairness. As George

Stigler (1981, p. 176) wrote: "[When] self-interest and ethical values with wide verbal allegiance are in conflict, much of the time, most the time in fact, self-interest-theory . . . will win."

Ultimatum games are a way to test whether Stigler was right. A Responder who rejects an offer of $2 from a $10 sum puts a $2 price tag on how much she dislikes being treated unfairly ("negative reciprocity").

The emotional reaction to unfairness which is highlighted by the ultimatum game can work at many levels. A reviewer suggested a dramatic illustration from political history. At the Federal Convention in 1787 in Philadelphia, delegates from the original thirteen American states debated how to treat new states that would join later, as western lands were annexed. Gouverneur Morris argued they should be admitted as second-rate states, so they could not outvote the original thirteen. George Mason argued that offering second-rate status would be like offering somebody an unfair portion, and the western states might reject it. He said, "They [new western states] will have the same pride and other passions which we have, and will either not unite with or will speedily revolt from the Union, if they are not in all respects placed on an equal footing with their brethren" (Farrand, 1966, pp. 578–79). Mason's argument that western states might reject an unequal offer (and the moral appeal of equal treatment, independent of the threat of rejection) eventually won over the delegates. (If he hadn't, I might be writing this in the great *country* of California, under the political aegis of California President Arnold Schwarzenegger and Vice-President Shaquille O'Neal, rather than under the rule of George Bush, the minority choice of voters in the impoverished neighbor country America, to our east.)

This chapter discusses several other games that measure aspects of social preference (see also Bolton, 1998; Sobel, 2001). Dictator games are ultimatum games with the Responder's ability to reject the offer removed. Dictator games establish whether Proposers in ultimatum games make generous offers because they fear rejection or because they are purely altruistic. (The answer is mostly fear and a little altruism.) Trust games are dictator games with an initial investment by an Investor that determines how much the dictator, or Trustee, has to allocate. The Investor "trusts" that the Trustee will give back enough to make her initial trust worthwhile. Trust games are simple models of contracting with moral hazard and no contractual enforcement. The amount of investment measures trust; repayment measures trustworthiness. The "centipede" game (see Chapter 5) is a trust game with several stages. A multiplayer trust game is "gift exchange" in labor markets: Firms offer wages to workers who accept offers and choose effort levels. Effort is costly to workers and valuable to firms but cannot be enforced by firms, so firms must trust workers to work hard. Experimentation with ultimatum, dictator, and trust games has exploded recently. These games are popular

***Table 2.1.*** *Prisoners' dilemma*

| | Cooperate | Defect |
|---|---|---|
| Cooperate | H,H | S,T |
| Defect | T,S | L,L |

*Note:* Assumes $T > H > L > S$.

because they model key features of economic situations and experiments are easy to run.

Two other important games are prisoners' dilemma (PD) and public goods games. Although these games have been studied in literally thousands of experiments, I will say only a little about them because the results are well known and nicely summarized in many other sources (Davis and Holt, 1993; Colman, 1995, Chapter 7; Ledyard, 1995; Sally, 1995).

Table 2.1 shows payoffs in a typical PD. Mutual cooperation provides payoffs of H for each player, which is better than the L payoff from mutual defection. However, if the other player cooperates, a defector earns the T(emptation) payoff, which is better than reciprocating and earning only H (since $T > H$ in a PD). A player who cooperates against a defector earns the S(ucker) payoff, which is less than earning L from defecting. Since $T > H$ and $L > S$, both players prefer to defect whether the other player cooperates or not. Mutual defection is the only Nash equilibrium but it is Pareto dominated (worse for both players) than mutual cooperation.

In public goods games, each of $N$ players can invest resources $c_i$ from their endowment $e_i$ in a public good that is shared by everyone and has a total per-unit value of $m$.[1] Player $i$ earns $e_i - c_i + m(\sum_k c_k)/N$. Assuming $m < 1/N$, the payoff-maximizing outcome is to contribute nothing ($c_i = 0$). If everyone contributed, however, the players would collectively earn the most.

PD and public goods games are models of situations such as pollution of the environment, in which one player's action imposes a harmful "externality" on innocent parties (cooperation corresponds to voluntarily limiting pollution), villagers sharing a depletable resource such as river water, and production of a public utility such as a school or irrigation system that non-contributing "free riders" cannot easily be excluded from sharing (see Ostrom, 2000). Low rates of voluntary cooperation and contribution in these games can be remedied by institutional arrangements such as government

[1]Assume $m < 1$ so that a player does not benefit enough personally to contribute for private gain, and $mN > 1$ so all players contributing is Pareto improving.

taxation (which forces free riders to pay up), or informal mechanisms such as yelling at people who throw trash out of their cars (ostracism). And when players in PD and public goods games are matched together repeatedly, it can be an equilibrium for all players to cooperate until one player defects.

In experiments, subjects cooperate in one-shot PD games about half the time and contribute about half their endowments in public goods game (although there is wide dispersion; most subjects contribute either all or nothing—see Sally, 1995; Ledyard, 1995). Changes in monetary payoffs have predictable effects: Lowering $T$ and raising $S$ increase cooperation in the PD; and raising the marginal return $m$ raises public good contribution. Preplay communication, which should have no effect in theory, is the non-payoff variable that raises the rate of cooperation by the most.

When the games are repeated with random "stranger" rematching, cooperation and contribution dwindle to a small core of persistent contributors. Figure 2.1 shows a typical pattern of average contributions over time from Fehr and Gächter (2000c) in stranger treatments. The figure also shows the strong effect of punishment. If free-riding is prevalent because of self-interest, then players should not incur a private cost to punish free riders (they should free ride on punishment by others, a "second-order" free-riding problem). This prediction is wrong: Yamagishi (1986) and Fehr and Gächter (2000c) showed that costly punishment works very well, raising steady-state contributions to more than half of endowments, as shown in Figure 2.1. (In partner protocols with rematching, contributions are nearly 100 percent. Then the punishment mechanism is "free" because, when contributions are that high, nobody punishes.)

Players who contribute are more likely to say they expect others to contribute than free riders are. This correlation between beliefs and choices implies that cooperation is conditional or reciprocal. Contributions are also higher when the same players are paired together in partner protocols (e.g., Andreoni, 1993), which is consistent with "folk theorems" about the efficiency gains from repeated play.

PD and public goods games are important in economic life, but they are blunt tools for guiding theories of social preference. These games cannot distinguish between players who are altruistic and players who match expected cooperation. Nor can they distinguish between players who are self-interested and those who have reciprocal preferences but pessimistically think others will free ride. The other games described in this chapter are sharper tools for making these sorts of distinctions than PD and public goods games.

Before proceeding, it is crucial to emphasize again (as Weibull, 2000, has) that evidence of public good contribution, dictator allocation, ultimatum rejection, and trust repayment does *not* falsify game theory, per se. Games lead to utilities over allocations, and one player's concern for how

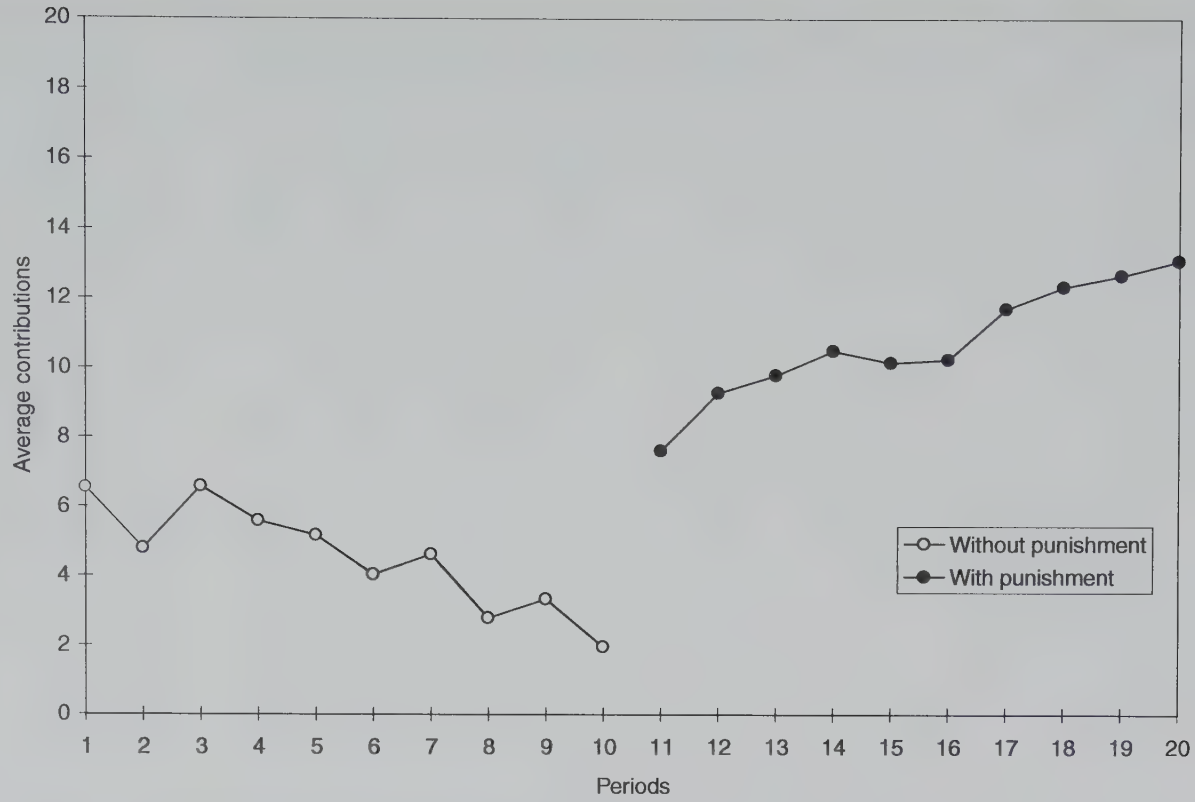



***Figure 2.1.*** *Public good contributions before and after punishment. Source: Fehr and Gächter (2000c), p. 989, Figure 3B; reproduced with permission of the American Economic Association.*

much another player earns (whether positive or negative) can certainly affect her utility. In experiments, however, games are played in money. Since we cannot easily measure or control players' preferences over how much others earn, we always end up testing a joint hypothesis of game-theoretic behavior coupled with some assumption about utilities over money outcomes.

Offering $4 out of $10 could be an equilibrium offer in an ultimatum game, if the Proposer correctly believes that the Responder has a strong dispreference for unequal allocations and will reject a lower offer. But offering $4 could also be a disequilibrium offer, if the Proposer's belief about the Responder's likely behavior is wrong.

Thus, the trick in explaining these data from simple bargaining and public goods games is to seek a parsimonious specification of how observable payments (such as dollars paid to each individual subject) map into an individual's "social preferences," then test the *joint* hypothesis that players have a particular kind of social preferences and play game-theoretically. Throughout the chapter, you will see that the joint hypothesis of game-theoretic behavior *and* social preferences that value only one's own payment (what is often, casually but imprecisely, called "the" game-theoretic prediction) is easily rejected. Then the interesting question is, "Are the rejections due to the pure self-interest part of the joint hypothesis, or to the game-theoretic reasoning part (or both)?"

The answer seems to be "both"; so both parts of the joint hypothesis are in need of repair. Deciding whether to accept an ultimatum offer requires no strategic thinking (it is simply a choice), so rejections clearly indict the self-interest assumption as the culprit. At the same time, direct evidence of how people think during the game (in Chapter 4) and of limited strategic thinking in constant-sum dominance-solvable games (in Chapter 5) shows that game-theoretic reasoning is limited even when self-interest is a reasonable assumption. So is game theory dead? Of course not; it is simply being renovated, or generalized in a precise way. Simple parsimonious models of both social preferences (see Section 2.8 in this chapter) and limited strategic thinking (e.g., Chapter 5 and Camerer, Ho, and Chong, 2001) have already emerged and should continue to be a hot topic of research.

## 2.1 Ultimatum and Dictator Games: Basic Results

In a typical ultimatum experiment, subjects are paired with anonymous others, and a Proposer makes an offer that the Responder then accepts or rejects. Two common variants on this baseline design either repeat the game (rematching with a new player each time) or ask the Responder to state a minimum acceptable offer (MAO) rather than simply decide whether to accept a specific offer. The MAO method has the huge advantage of

measuring likely reactions to all possible offers, which is important if the most interesting offers (such as very low ones) are rare. For some reason, economists are generally reluctant to use the MAO method (even if very little is learned about rejection behavior by using the specific-offer method).[2]

Tables 2.2 and 2.3 compile statistics from many studies of ultimatum games. These studies used specific offers in one-shot games, unless noted otherwise. Each line of the tables lists a study, the amount being divided, and the number of pairs of subjects. The data are relative frequencies of offers in each percentage interval (Table 2.2) and the relative frequency of rejections conditional on each percentage offer (Table 2.3). The median offer percentage is printed in italics. Mean offers and overall rejection rates are shown in the rightmost columns of each table. (Percentages are fractions of the total amount being divided unless stated otherwise.) Significant differences across conditions within a study are noted by lettered "significance codes." Conditions with the same letter are not significantly different; conditions with different letters are significantly different at $p \leq 0.05$.

The results reported in the tables are very regular. Modal and median ultimatum offers are usually 40–50 percent and means are 30–40 percent. There are hardly any offers in the outlying categories of 0, 1–10, and the hyper-fair category 51–100. Offers of 40–50 percent are rarely rejected. Offers below 20 percent or so are rejected about half the time.

It is useful to distinguish the emotions or reasons that cause Responders to reject behavior (call it "anger") from the emotion A might feel when B does something unfair to a third party, C (call it "indignation"). Anger is more personal, and often motivates an aggrieved party to administer justice herself. An indignant A is cooler and is more likely to be happy if B is punished in some other way. (These delicate emotions are important in shaping proceedings such as war tribunals, as in the Nuremberg trials or the South African Truth and Reconciliation Commission. Indignation by parties who were not directly harmed, but are appalled that others were, is important but probably less powerful a force than personal anger.) Indignation is cooler. In recent work, Ernst Fehr and colleagues have been exploring "third-party punishment" games in which A can spend money to punish B for treating C unfairly.

Generous offers could come about because Proposers are fair-minded or because Proposers are afraid of having low offers rejected (or both). The two explanations can be easily separated in a dictator game, which removes the Responder's ability to reject offers. If Proposers offer positive amounts

[2] There is a vague sense that Responders demand more when stating MAOs (see Weber and Camerer, 2001), but this difference is neither well established nor (if true) well understood.

***Table 2.2.*** *Frequencies of ultimatum offers*

| Reference<br>Experimental condition | Amount($) | No. of pairs | Offer frequencies (percent offered)<br>0 | 1–10 | 11–20 | 21–30 | 31–40 | 41–50 | 51–60 | 61–100 | Mean offer | Signif. code | Comments |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Bolton and Zwick (1995)* | | | | | | | | | | | | | |
| Cardinal ultimatum | 4 | 20 | < 0.45 > | | *0.25* | 0.05 | 0.00 | 0.25 | | | 0.24 | a | Offers imputed from |
| Double blind (ZK) | 4 | 20 | *<0.50>* | | *0.25* | 0.00 | 0.00 | 0.25 | | | 0.22 | a | series of games |
| *Cameron (1999)* | | | | | | | | | | | | | |
| Indonesian rupiah | 5K | 101 | 0.08 | 0.02 | 0.12 | 0.06 | 0.20 | *0.38* | 0.03 | 0.11 | 0.42 | a | "Problems" excluded |
| Indonesian rupiah | 40K | 35 | | | | 0.17 | 0.17 | *0.63* | | 0.03 | 0.45 | b | |
| Indonesian rupiah | 200K | 37 | 0.03 | 0.03 | 0.03 | 0.08 | 0.24 | *0.57* | 0.03 | | 0.42 | a | |
| *Croson (1996)* | | | | | | | | | | | | | |
| Informed | 10 | 26 | 0.00 | 0.04 | 0.04 | 0.04 | 0.25 | *0.57* | 0.07 | | 0.45 | a | Data from working |
| Uninformed | 10 | 28 | 0.00 | 0.12 | 0.15 | 0.15 | *0.23* | 0.31 | | 0.04 | 0.36 | b | paper |
| *Eckel and Grossman (2001)* | | | | | | | | | | | | | |
| Women | 5 | 96 | 0.01 | 0.01 | 0.11 | 0.14 | *0.52* | 0.21 | 0.01 | 0.01 | 0.41 | a | Numbers estimated |
| Men | 5 | 95 | 0.01 | 0.04 | 0.12 | 0.22 | *0.45* | 0.18 | 0.01 | | 0.39 | a | from figures |
| *Hoffman, McCabe, Shachat, and Smith (1994); Hoffman, McCabe, and Smith (1996a)* | | | | | | | | | | | | | |
| FHSS replication | 10 | 24 | | | | 0.13 | *0.38* | *0.50* | | | 0.44 | a | |
| Contest | 10 | 24 | | 0.08 | 0.25 | *0.25* | 0.33 | 0.08 | | | 0.31 | b | |
| FHSS replication | 100 | 27 | 0.04 | | | 0.11 | 0.26 | *0.52* | 0.07 | | 0.44 | a | |
| Contest | 100 | 23 | | 0.17 | 0.26 | *0.26* | 0.22 | 0.11 | | | 0.29 | b | |
| *Güth, Schmittberger, and Schwarze (1982)* | | | | | | | | | | | | | |
| Naive | 4–10 | 21 | 0.10 | | 0.14 | 0.10 | 0.24 | 0.43 | | | 0.37 | | Payment in DM |
| Experienced | 4–10 | 21 | | 0.05 | 0.11 | *0.29* | 0.24 | 0.24 | | | 0.33 | | 9 and 12 games experience |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Roth, Prasnikar, Okuno-Fujiwara, and Zamir (1991)* | | | | | | | | | | | | | |
| Pittsburgh, Round 1 | 10 | 27 | | | 0.04 | 0.11 | 0.22 | *0.56* | 0.04 | 0.04 | 0.47 | | |
| Pittsburgh, Round 1 | 30 | 10 | | | | | | *0.80* | 0.20 | | 0.52 | | |
| Yugoslavia, Round 1 | 400K | 30 | | | 0.03 | 0.07 | 0.13 | *0.73* | 0.03 | | 0.46 | | Payment in dinars |
| Japan, Round 1 | 2000 | 29 | | 0.07 | 0.07 | 0.14 | 0.14 | *0.41* | 0.07 | 0.07 | 0.42 | | Payment in yen |
| Israel, Round 1 | 20 | 30 | | 0.17 | 0.07 | 0.10 | *0.20* | 0.43 | 0.03 | | 0.37 | | Payment in shekels |
| Pittsburgh, Round 10 | 10 | 27 | | | 0.04 | | 0.33 | *0.63* | | | 0.46 | a | |
| Pittsburgh, Round 10 | 30 | 10 | | | | 0.10 | | 0.80 | 0.10 | | 0.49 | a | |
| Yugoslavia, Round 10 | 400K | 30 | | | | 0.03 | 0.27 | *0.70* | | | 0.47 | a | Payment in dinars |
| Japan, Round 10 | 2000 | 29 | | | | 0.17 | *0.34* | 0.48 | | | 0.43 | b | Payment yen |
| Israel, Round 10 | 20 | 30 | | 0.03 | 0.13 | 0.20 | *0.57* | 0.07 | | | 0.35 | c | Payment in shekels |
| *Forsythe, Horowitz, Savin, and Sefton (1994)* | | | | | | | | | | | | | |
| With pay | 10 | 24 | | | 0.04 | 0.04 | 0.17 | *0.71* | 0.04 | | 0.47 | a | |
| With pay | 5 | 43 | | | 0.09 | 0.02 | 0.23 | *0.53* | 0.11 | | 0.45 | a | |
| Without pay | 5 | 48 | 0.04 | | 0.06 | | 0.31 | *0.48* | 0.08 | 0.02 | 0.44 | a | Two sessions different |
| *Harrison and McCabe (1996b)* | | | | | | | | | | | | | |
| U1 (public display), period 20 | | 16 | | | 0.06 | | 0.19 | *0.75* | | | 0.44 | a | MAOs; U3 reports |
| U1 (public display), period 20 | | 16 | | 0.31 | *0.50* | 0.19 | | | | | 0.13 | b | human data only |
| U3 (computer), period 1 | 20 | 16 | | | 0.12 | 0.19 | *0.56* | 0.06 | 0.06 | | 0.46 | a | |
| UC (computer), period 15 | 20 | 16 | | | 0.19 | *0.81* | | | | | 0.14 | c | |
| *Larrick and Blount (1997)* | | | | | | | | | | | | | |
| Ultimatum (control) | 7 | 51 | 0.02 | 0.12 | 0.10 | 0.13 | 0.04 | *0.57* | 0.02 | | 0.38 | a | 1/3 paid; nos. from |
| Claiming language | 7 | 54 | 0.06 | 0.08 | 0.02 | 0.13 | 0.02 | *0.67* | 0.02 | | 0.41 | a | figures (study 1); MAOs |
| *Rapoport, Sundali, and Potter (1996)* | | | $0–0.99 | 1–1.99 | 2–2.99 | 3–3.99 | 4–4.99 | 5–6.99 | 7–8.99 | 9–max | | | Proposers informed, |
| Amount uniform [0,30] | [0,30] | 10 | 0.13 | 0.08 | 0.16 | 0.10 | *0.08* | 0.25 | 0.10 | 0.11 | 0.31 | a | make $ offers |
| Amount uniform [5,25] | [5,25] | 10 | 0.04 | 0.01 | 0.08 | 0.16 | 0.17 | *0.23* | 0.18 | 0.12 | 0.38 | b | 2 ten-period phases, |
| Amount uniform [10,20] | [10,20] | 10 | | 0.01 | | 0.06 | 0.16 | *0.52* | 0.18 | 0.07 | 0.41 | c | roles switched |

*(continued)*

***Table 2.2.*** *(continued)*

| Reference<br>Experimental condition | Amount ($) | No. of pairs | Offer frequencies (percent offered)<br>0 | 1–10 | 11–20 | 21–30 | 31–40 | 41–50 | 51–60 | 61–100 | Mean offer | Signif. code | Comments |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Rapoport, Sundali, and Seale (1996)* | | Class: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | | Proposers informed |
| Amount uniform [0,30] | [0,30] | 20 | 0.13 | 0.07 | 0.17 | 0.12 | *0.18* | 0.13 | 0.12 | 0.08 | 0.28 | a | |
| Amount uniform [5,25] | [5,25] | 20 | 0.04 | 0.09 | 0.15 | 0.19 | *0.12* | 0.17 | 0.19 | 0.04 | 0.34 | b | |
| Amount uniform [10,20] | [10,20] | 20 | 0.14 | 0.08 | 0.06 | 0.16 | *0.14* | 0.19 | 0.12 | 0.08 | 0.35 | b | |
| *Schotter, Weiss, and Zapater (1996)* | | | | | | | | | | | | | |
| One stage | 10 | 17 | | | 0.06 | 0.18 | 0.12 | *0.53* | 0.12 | | 0.45 | a | |
| First of two stages | 10 | 18 | | 0.17 | 0.17 | 0.17 | *0.06* | 0.17 | 0.25 | 0.06 | 0.41 | a | |
| *Slonim and Roth (1998)* | | | | | | | | | | | | | |
| Low stakes | 60 | 240 | | | 0.01 | 0.03 | 0.16 | *0.75* | 0.06 | | 0.45 | a | Payment in |
| Medium stakes | 300 | 330 | | 0.04 | | 0.07 | 0.20 | *0.66* | 0.07 | | 0.42 | a | Slovak crowns |
| High stakes | 1500 | 250 | | 0.01 | 0.06 | 0.04 | 0.12 | *0.69* | 0.07 | | 0.43 | a | |
| *List and Cherry (2000)* | | | | | | | | | | | | | |
| Low stakes | 20 | 290 | | | 0.28 | 0.10 | *0.17* | 0.36 | 0.09 | | 0.34 | a | Offers < 25 percent |
| High stakes | 400 | 270 | | | 0.27 | 0.17 | *0.17* | 0.34 | 0.04 | | 0.32 | a | in (11,20) interval |

*Note:* Figures in italic represent median offers.

***Table 2.3.*** *Frequencies of rejections in ultimatum games*

| Reference<br>Experimental condition | Amount($) | No. of pairs | Conditional rejection frequencies (percent offered)<br>0 | 1–10 | 11–20 | 21–30 | 31–40 | 41–50 | 51–60 | 61–100 | Rejection rate | Comments |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Bolton and Zwick (1995)* | | | | | | | | | | | | |
| Cardinal ultimatum | 4 | 20 | < 1.00 > | | 0.78 | 0.57 | 0.12 | 0.08 | | | 0.38 | Offers imputed from |
| Double blind (ZK) | 4 | 20 | < 1.00 > | | 0.70 | 0.07 | 0.07 | 0.13 | | | 0.30 | series of games |
| *Cameron (1999)* | | | | | | | | | | | | |
| Indonesian rupiah | 5K | 101 | 1.00 | 1.00 | 0.75 | 1.00 | 0.08 | 0.03 | 0.00 | 0.00 | 0.17 | "Problems" excluded |
| Indonesian rupiah | 40K | 35 | | | | 0.40 | 0.17 | 0.00 | | 0.00 | 0.09 | |
| Indonesian rupiah | 200K | 37 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.05 | | | 0.12 | |
| *Croson (1996)* | | | | | | | | | | | | |
| Informed | 10 | 26 | | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | | 0.07 | Data from working |
| Uninformed | 10 | 28 | | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | | 0.00 | 0.04 | paper |
| *Eckel and Grossman (2001)* | | | | | | | | | | | | |
| Women | 5 | 96 | 1.00 | 0.50 | 0.35 | 0.23 | 0.03 | 0.00 | 0.00 | 0.00 | 0.10 | Numbers estimated |
| Men | 5 | 95 | 0.00 | 1.00 | 0.50 | 0.40 | 0.05 | 0.00 | 0.00 | | 0.14 | from figures |
| *Hoffman, McCabe, Shachat, and Smith (1994); Hoffman, McCabe, and Smith (1996a)* | | | | | | | | | | | | |
| FHSS replication | 10 | 24 | | | | 0.33 | 0.11 | 0.00 | | | 0.08 | |
| Contest | 10 | 24 | | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | | | 0.13 | |
| FHSS replication | 100 | 27 | | | | 0.00 | 0.14 | 0.00 | 0.00 | | 0.04 | |
| Contest | 100 | 23 | | 0.75 | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 | | 0.21 | |
| *Güth, Schmittberger, and Schwarze (1982)* | | | | | | | | | | | | |
| Naive | 4–10 | 21 | 0.50 | | 0.33 | 0.00 | 0.00 | 0.00 | | | 0.10 | Payment in DM |
| Experienced | 4–10 | 21 | | 1 | 0.50 | 0.33 | 0.00 | 0.20 | | | 0.25 | 9 and 12 games experience |

*(continued)*

*Table 2.3.* *(continued)*

| Reference Experimental condition | Amount($) | No. of pairs | Conditional rejection frequencies (percent offered) 0 | 1–10 | 11–20 | 21–30 | 31–40 | 41–50 | 51–60 | 61–100 | Rejection rate | Comments |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Roth, Prasnikar, Okuno-Fujiwara, and Zamir (1991)* | | | | | | | | | | | | |
| Pittsburgh, Round 1 | 10 | 27 | | | 0.00 | 0.67 | 0.17 | 0.07 | 0.00 | 1.00 | 0.19 | |
| Pittsburgh, Round 1 | 30 | 10 | | | | | | 0.25 | 0.00 | | 0.20 | |
| Yugoslavia, Round 1 | 400K | 30 | | | 1.00 | 0.50 | 0.25 | 0.23 | 0.00 | | 0.27 | Payment in dinars |
| Japan, Round 1 | 2000 | 29 | | 0.50 | 0.50 | 0.50 | 0.25 | 0.08 | 0.50 | 0.00 | 0.24 | Payment in yen |
| Israel, Round 1 | 20 | 30 | | 0.80 | 0.50 | 0.33 | 0.00 | 0.15 | 0.00 | | 0.27 | Payment in shekels |
| Pittsburgh, Round 10 | 10 | 27 | | | 1.00 | | 0.22 | 0.12 | | | 0.19 | |
| Pittsburgh, Round 10 | 30 | 10 | | | | 0.00 | | 0.13 | 0.00 | | 0.10 | |
| Yugoslavia, Round 10 | 400K | 30 | | | | 1.00 | 0.63 | 0.05 | | | 0.24 | Payment in dinars |
| Japan, Round 10 | 2000 | 29 | | | | 0.20 | 0.10 | 0.14 | | | 0.14 | Payment yen |
| Israel, Round 10 | 20 | 30 | | 0.00 | 0.25 | 0.17 | 0.12 | 0.00 | | | 0.13 | Payment in shekels |
| *Forsythe, Horowitz, Savin, and Sefton (1994)* | | | | | | | | | | | | |
| With pay | 10 | 24 | | | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | | 0.04 | |
| With pay | 5 | 43 | | | 0.50 | 0.00 | 0.10 | 0.00 | 0.00 | | 0.07 | |
| Without pay | 5 | 48 | 1.00 | | 0.67 | | 0.27 | 0.00 | 0.00 | 0.00 | 0.17 | Two sessions different |
| *Hoffman, McCabe, Shachat, and Smith (1994)* | | | | | | | | | | | | |
| FHSS replication | 10 | 24 | | | | 0.33 | 0.11 | 0.00 | | | 0.08 | |
| FHSS and Contest | 10 | 24 | | | 0.00 | 0.00 | 0.00 | 0.00 | | | 0.00 | |
| Exchange | 10 | 24 | | | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | | 0.08 | |
| Exchange and Contest | 10 | 24 | | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | | | 0.13 | |
| *Harrison and McCabe (1996b)* | | | | | | | | | | | | |
| U1 (public display), period 1 | 20 | 16 | 1.00 | 0.87 | 0.74 | 0.61 | 0.36 | 0.00 | 0.00 | 0.00 | 0.11 | MAOs; U3 reports |
| U1 (public display), period 15 | 20 | 16 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.31 | human data only |
| U3 (computer), period 1 | 20 | 16 | 1.00 | 1.00 | 1.00 | 0.94 | 0.63 | 0.51 | 0.00 | 0.00 | 0.68 | |
| UC (computer), period 15 | 20 | 16 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Larrick and Blount (1997)* | | | | | | | | | | | | |
| Ultimatum (control) | 7 | 51 | 0.98 | 0.67 | 0.58 | 0.46 | 0.37 | 0.03 | 0.00 | 0.00 | 0.30 | 1/3 paid; nos. from |
| Claiming language | 7 | 54 | 0.70 | 0.48 | 0.37 | 0.26 | 0.26 | 0.04 | 0.00 | 0.00 | 0.15 | figures (study 1); MAOs |
| *Rapoport, Sundali, and Potter (1996)* | | | $:0-.99 | 1-1.99 | 2-2.99 | 3-3.99 | 4-4.99 | 5-6.99 | 7-8.99 | 9-max | | Proposers informed, |
| Amount uniform [0,30] | [0,30] | 10 | 0.68 | 0.40 | 0.19 | 0.20 | 0.06 | 0.00 | 0.00 | 0.05 | 0.18 | make $ offers |
| Amount uniform [5,25] | [5,25] | 10 | 0.86 | 0.40 | 0.31 | 0.18 | 0.09 | 0.02 | 0.00 | 0.00 | 0.12 | 2 ten-period phases, |
| Amount uniform [10,20] | [10,20] | 10 | | 1.00 | | 0.67 | 0.67 | 0.18 | 0.00 | 0.00 | 0.25 | roles switched |
| *Rapoport, Sundali, and Seale (1996)* | | class: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | Proposers informed |
| Amount uniform [0,30] | [0,30] | 20 | 0.29 | 0.22 | 0.09 | 0.18 | 0.01 | 0.16 | 0.15 | 0.30 | | |
| Amount uniform [5,25] | [5,25] | 20 | 0.65 | 0.22 | 0.12 | 0.06 | 0.10 | 0.04 | 0.11 | 0.13 | | |
| Amount uniform [10,20] | [10,20] | 20 | 0.55 | 0.39 | 0.22 | 0.15 | 0.09 | 0.07 | 0.02 | 0.06 | | |
| *Schotter, Weiss, and Zapater (1996)* | | | | | | | | | | | | |
| One stage | 10 | 17 | | | 1.00 | 0.33 | 0.50 | 0.00 | 0.00 | | 0.18 | |
| First of two stages | 10 | 18 | | 1.00 | 0.33 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.28 | |
| *Slonim and Roth (1998)* | | | | | | | | | | | | |
| Low stakes | 60 | 240 | | | 1.00 | 0.67 | 0.42 | 0.11 | 0.07 | | 0.18 | Payment in Slovak crowns |
| Medium stakes | 300 | 330 | | 0.85 | | 0.31 | 0.17 | 0.07 | 0.05 | | 0.16 | |
| High stakes | 1500 | 250 | | 0.50 | 0.50 | 0.58 | 0.07 | 0.03 | 0.00 | | 0.14 | |
| *List and Cherry (2000)* | | | | | | | | | | | | |
| Low stakes | $20 | 290 | | | 0.72 | 0.43 | 0.30 | 0.13 | 0.12 | | 0.35 | |
| High stakes | $400 | 270 | | | 0.55 | 0.28 | 0.17 | 0.08 | 0.00 | | 0.26 | |

in a dictator game, they are not payoff maximizing, which suggests some of their generosity in ultimatum games is altruistic rather than strategic.

In the first dictator game experiment, Kahneman, Knetsch, and Thaler (1986) gave subjects a choice between dictating an even split of $20 with another student, or an uneven ($18,$2) split favoring themselves. Their results are shown in Table 2.4, which compiles statistics from many dictator game experiments. Three-quarters chose the equal split ($10,$10).

Reaction to the striking dictator results got the literature off on the wrong interpretive foot. Many people thought the main question about the ultimatum findings was whether offers were fair or were strategic (merely avoiding rejection). But the tail that wags the proverbial dog is the *rejections by Responders*, which force Proposers to make generous offers. Forsythe et al. (1994) did the first thorough comparison of dictator and ultimatum results where dictators could offer any amount they wanted (rather than simply choosing one of two allocations). Their dictators show less generosity than Kahneman et al. reported, but the mean allocation is about 20 percent, showing that there is some pure altruism. The fact that dictator offers are much lower than Proposer offers in ultimatum games, but positive, shows that Proposers are being both strategic (avoiding more to avoid rejection) *and* altruistic.[3] Early results showed that average offers are close to the offer that maximizes expected payoffs given the actual pattern of rejections (e.g., Roth et al., 1991), which implies Proposers are simply being strategic. More sophisticated analyses then showed that actual offers are more generous than payoff-maximizing offers, even controlling for risk-aversion (Henrich et al., 2001; cf. Lin and Sunder, 2002). In a model allowing nonequilibrium beliefs (and learning), Costa-Gomes and Zauner (2001) found that Proposer beliefs were generally a little too pessimistic.

The many studies on ultimatum and dictator games have varied the conditions in the game or identity of subjects to explore a wide variety of issues. Variables fall into five categories. *Methodological* variables change how the experiment is conducted—stakes, anonymity, and repetition. *Demographic* variables measure how different groups of people behave. (Few

[3] I suspect that Proposers behave strategically in ultimatum games because they expect Responders to stick up for themselves, whereas they behave more fair-mindedly in dictator games because Recipients cannot stick up for themselves. This behavior could be codified in a theory of reciprocal fairness that includes responsibility. Define the last-moving player who affects player $i$'s payoff as the only one 'responsible' for $i$. If that responsible player is not $i$ then she must take some care to treat $i$ fairly; otherwise, she can treat $i$ neutrally and expect $i$ to be responsible for herself. This idea is exemplified by former Philadelphia mayor Frank Rizzo, who was asked whether he ever gave money to the city's many homeless beggars. Rizzo was a notorious law-and-order autocrat (for example, the number of civilians shot by policemen fell dramatically after he retired as police chief), so I expected a gruff answer about how the homeless don't deserve handouts. Instead, Rizzo said he made a judgment about whether the person asking for money was capable of working (based on apparent physical or mental disability), and gave money to those beggars who appeared unable to work. Rizzo is providing social insurance.

***Table 2.4.*** *Allocations in dictator games*

| Reference<br>Experimental condition | $ | No. of pairs | Percent allocated to other person<br>0 | 1–10 | 11–20 | 21–30 | 31–40 | 41–50 | 51–60 | 61–70 | 71–90 | 91–100 | Mean | Signif. code | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Frey and Bohnet (1997)* | | | | | | | | | | | | | | | |
| One-way ID | 13 | 18 | 0.11 | | 0.06 | 0.17 | *0.22* | 0.44 | | | | | 0.35 | a | Swiss francs |
| One-way ID + info. | 13 | 25 | | 0.04 | 0.04 | 0.04 | 0.20 | *0.28* | 0.12 | 0.12 | 0.04 | 0.12 | 0.52 | b | |
| *Bolton, Katok, and Zwick (1998)* | | | | | | | | | | | | | | | |
| 1 Game 2 Card | 10 | 28 | 0.93 | na | na | na | na | 0.07 | na | na | na | na | | | |
| Kindness | 10 | 28 | na | na | na | na | 0.89 | 0.11 | na | na | na | na | na | | |
| 10 Game 6 Card | 1 | 25 | 0.40 | 0.04 | *0.36* | 0.16 | | 0.04 | na | na | na | na | 0.16 | a | 10-game average |
| 10 Game 2 Card | 1 | 25 | 0.40 | 0.08 | *0.24* | 0.20 | | 0.08 | na | na | na | na | 0.20 | a | 10-game average |
| Anonymity | 10 | 33 | 0.37 | *0.18* | 0.15 | 0.03 | 0.12 | 0.09 | | 0.03 | | | 0.17 | a | |
| 1 Game 6 Card | 10 | 27 | *0.52* | 0.15 | 0.07 | 0.07 | | 0.15 | na | na | na | na | 0.13 | a | |
| *Cason and Mui (1998)* | | | | | | | | | | | | | | | |
| Round 1 | 40 | 40 | 0.38 | 0.05 | 0.05 | *0.15* | 0.16 | 0.19 | | | 0.03 | | 0.23 | | |
| Round 2 | 40 | 40 | 0.28 | 0.16 | 0.05 | *0.05* | 0.05 | 0.12 | 0.12 | 0.05 | 0.13 | | 0.31 | | |
| *Forsythe, Horowitz, Savin, and Sefton (1994)* | | | | | | | | | | | | | | | |
| With pay | 10 | 24 | 0.21 | 0.17 | *0.13* | 0.29 | | 0.21 | | | | | 0.24 | b | |
| Without pay | 5 | 45 | 0.14 | | 0.11 | | *0.26* | 0.47 | | | 0.02 | | 0.38 | a | |
| With pay | 5 | 45 | 0.35 | | *0.28* | 0.05 | 0.09 | 0.18 | 0.05 | | | | 0.23 | b | |
| *Frey and Bohnet (1995)* | | | | | | | | | | | | | | | |
| Recipient ID'd | 13 | 39 | 0.28 | 0.08 | 0.03 | 0.10 | *0.18* | 0.30 | 0.03 | | | | 0.26 | a | Swiss francs |
| Mutual ID | 13 | 28 | | | 0.07 | | | *0.082* | | 0.04 | | 0.07 | 0.50 | b | |
| Mutual ID + communication | 13 | 17 | 0.06 | 0.06 | | 0.12 | 0.05 | *0.41* | 0.12 | | | 0.18 | 0.48 | b | |

*(continued)*

*Table 2.4.* *(continued)*

| Reference<br>Experimental condition | $ | No. of pairs | Percent allocated to other person<br>0 | 1–10 | 11–20 | 21–30 | 31–40 | 41–50 | 51–60 | 61–70 | 71–90 | 91–100 | Mean | Signif. code | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Frohlich and Oppenheimer (1997)* | | | | | | | | | | | | | | | |
| Canada | 10 | 22 | 0.34 | *0.18* | | 0.05 | 0.09 | 0.23 | | | | 0.11 | 0.27 | a | |
| United States | 10 | 19 | 0.47 | *0.20* | 0.05 | | | 0.26 | | | | | 0.16 | a | |
| *Grossman and Eckel (1993)* | | | | | | | | | | | | | | | |
| Double blind 1 | 10 | 12 | *0.58* | | 0.08 | 0.08 | 0.17 | 0.08 | | | | | 0.15 | | |
| Red Cross recipient | 10 | 48 | 0.27 | 0.10 | *0.23* | | 0.08 | 0.17 | | | 0.04 | 0.10 | 0.31 | | |
| *Hoffman, McCabe, Shachat, and Smith (1994)* | | | | | | | | | | | | | | | |
| Exchange labels | 10 | 24 | 0.21 | 0.04 | 0.04 | *0.42* | 0.17 | 0.12 | | | | | 0.27 | a | |
| Contest and exchange | 10 | 24 | 0.42 | *0.17* | 0.21 | 0.17 | 0.04 | | | | | | 0.13 | b | |
| Double blind 1 | 10 | 36 | *0.64* | 0.19 | 0.06 | 0.03 | | 0.06 | | | 0.03 | | 0.10 | c | |
| Double blind 2 | 10 | 41 | *0.59* | 0.20 | 0.02 | 0.07 | 0.02 | 0.10 | | | | | 0.10 | c | |
| *Hoffman, McCabe, and Smith (1996b)* | | | | | | | | | | | | | | | |
| FHSS replication | 10 | 28 | 0.18 | 0.18 | 0.07 | *0.18* | 0.07 | 0.25 | | < 0.07 > | | | 0.24 | | |
| FHSS variation | 10 | 28 | 0.43 | | *0.11* | 0.14 | 0.11 | 0.18 | | < 0.04 > | | | 0.20 | | |
| Single blind 1 | 10 | 37 | 0.41 | *0.27* | 0.11 | 0.05 | 0.03 | 0.14 | | | | | 0.15 | | |
| Single blind 2 (dec. form) | 10 | 43 | 0.42 | *0.21* | 0.12 | 0.05 | 0.05 | 0.09 | | < 0.07 > | | | 0.13 | | |
| *Kahneman, Knetsch, and Thaler (1990)* | | | | | | | | | | | | | | | |
| Limited choice | 20 | 161 | na | 0.24 | na | na | na | 0.76 | na | na | na | na | na | | 10 percent paid |
| *Schotter, Weiss, and Zapater (1996)* | | | | | | | | | | | | | | | |
| One stage (control) | 10 | 16 | 0.13 | 0.06 | | 0.25 | *0.06* | *0.44* | | | | 0.06 | 0.39 | a | |
| 1st of two stages | 10 | 16 | 0.31 | | *0.19* | *0.31* | 0.13 | | | 0.06 | | | 0.23 | b | |

demographic effects have proved to be large or replicable, although there are intriguing effects of race, age, and beauty.) *Culture* seems to be important when sampled broadly: In simple societies with more "market integration," ultimatum offers are closer to even splits. *Descriptive* variables change the description of the game but not its structure. *Structural* variables change the game by adding moves. Methodological, demographic, and descriptive variables have proved to have modest effects that are often not robust across studies. Cultural and structural variables have bigger effects and are more helpful for building social preference theories.

## 2.2 Methodological Variables

Methodological variables were vigorously studied as interest in the ultimatum games first reported in 1982 caught fire several years later.

### *2.2.1 Repetition*

Several experiments have used stationary replication to see whether repeating simple bargaining games matters. Roth et al. (1991), Bolton and Zwick (1995), Knez and Camerer (1995), Slonim and Roth (1998), and List and Cherry (2000) repeated ultimatum games using stranger-matching. Bolton and Zwick did not observe an experience effect; the other studies show a slight (usually insignificant) tendency for offers and rejections to fall over time.

Subjects may adjust offers over time more strongly when they know what all other subjects have done, or when playing with programmed subjects with unusual tastes (e.g., pure self-interest). Harrison and McCabe (1996b) measured these effects. Providing information about the offers and MAOs of all other subjects lowered offers and MAOs drop to around 15 percent by period 15.[4] Either Responders quit punishing unfair Proposers if they see that many others are not doing so, or their perceptions of what is fair are shaped by what others are doing. In another condition, when data from sixteen human subjects were reported to subjects along with sixteen additional random offers and MAOs of 1–14 percent, human offers and MAOs fell substantially over time.

Taken together, these studies show only a small effect of experience (lowering offers and rejections) unless the population is 'seeded' with self-interested computerized players. Note that a drop in rejections could be

[4] See Harrison and McCabe's (1992) working paper.

learning, or it could be temporary satiation of a taste for revenge. If rejections are emotional expressions of distaste for being treated unfairly, those expressed tastes might temporarily satiate like other tastes with a visceral component (such as food, exercise, or sex). For example, Aristotle opined that "men become calm when they have spent their anger on someone else." And, in trials of collaborators in German-occupied countries after 1945, those who were tried later generally received milder sentences (holding severity of the crime constant). Workers in less-developed countries speak of "donor fatigue," when charitable impulses are worn out by overwhelming solicitation by the poor.

A very easy way to distinguish satiation from learning is to "restart" the experiment after a break of a day or a week. If subjects stopped rejecting in one session because they got tired of expressing their anger, the frequency of rejections should rise again after the break. If they stopped rejecting because they learned not to, the break should have little effect.

### *2.2.2 Methodology: Stakes*

If I had a dollar for every time an economist claimed that raising the stakes would drive ultimatum behavior toward self-interest, I'd have a private jet on standby all day. Many experimental studies *have* raised stakes (see Camerer and Hogarth, 1999).[5] In simple tasks such as ultimatum games, paying extra usually does not make much difference in how hard players think because the task is easy. But higher stakes might change the relative weight players put on their own payoffs and the payoffs of others. In fact, most sensible theories predict that, as stakes rise, the *amount* that Responders will reject goes up but the *percentage* they will reject goes down. (That is, they are more likely to reject $5 out of $50 than out of $10, and are more likely to accept 10 percent of a $50 pie than of a $10 pie.)

Studies show some stakes effect of this sort, but the effects are surprisingly weak. The earliest studies in the United States show no significant effect on rejection rates (Roth et al., 1991; Forsythe et al., 1994; Hoffman, McCabe, and Smith, 1996a; Straub and Murnighan, 1995). (These studies are handicapped, however, because the specific-offer method is used and low offers are rare, so the statistical power to detect changes in rejection rates is low.)

[5] Across many different experiments, the largest effect of raising incentives comes when subjects are paid some performance-related incentive, compared with being paid nothing, and when the task is neither so difficult that thinking harder doesn't help nor so easy that very little thinking is required (for both extremes the performance improvement from extra payment is low). Once subjects are paid some incentive, multiplying the stake by two or ten makes little difference for average responses. However, paying more generally reduces outliers and shrinks variance.

Inventive studies have been done in foreign countries where modest stakes (by American standards) have large purchasing power. Cameron (1999) did the first study in Indonesia. Her stakes were 5K, 40K, and 200K rupiah (about one day's to one month's wages); stakes made little difference. In the Slovak Republic, Slonim and Roth (1998) found significantly fewer rejections in medium- and high-stakes conditions pooling ten rounds of play (matching with no repetition in matching). List and Cherry (2000) did a clever high-stakes experiment in Florida (which is considered a foreign country by Californians) with an important twist: Subjects who answered more general knowledge questions correctly "earned" the right to make offers from a $400 pie rather than a measly $20 pie. List and Cherry conjectured, correctly, that entitlement would generate more low offers and therefore more statistical power to detect any difference in rejections for the two pie sizes (and change over time). Indeed, rejection rates are a little lower for the $400 pie and decline modestly over time.

Taken together, these studies show that very large changes in stakes (up to several months' wages) have only a modest effect on rejections. Raising stakes also has little effect on Proposers' *offers*, presumably because aversion to costly rejection leads subjects to offer *closer* to 50 percent when stakes go up.[6] The frequent number of rejections of large dollar amounts is striking. In Hoffman, McCabe, and Smith (1996a) two subjects (out of six) rejected $30 offers out of $100. In List and Cherry (2000) a quarter of the subjects who were offered $100 out of $400 rejected it. It is tempting to conclude that these subjects were "confused," but this explanation is acceptable only if confusion is measured independently of whether a subject's behavior deviated from somebody's pet theory.

A different kind of incentive effect changes the "price" at which people indulge tastes for vengeance or altruism. Andreoni and Miller (2002) ran dictator games in which they multiplied the amount allocated to the Recipient. A multiplier greater than one models situations in which charitable contributions are "matched" by employers (or by the government through tax deduction). If Dictators have preferences over dollar allocations between themselves and Recipients, they should allocate (weakly) more when the multiplier is high. Their subjects were endowed with forty to one hundred tokens (worth $0.10). Tokens could be "held" for a value from 1 to 3 to the Dictator, or "passed," giving a value of 1 to 3 for the Recipient. By varying the "price" of altruism—the ratio of token values—they are able to classify subjects into three categories (and forecast contributions in other public

[6] At a conference at Penn years ago, several of us discussed how much we'd offer from a $1 million stake. As I recall, Al Roth said he'd offer something like $400,000, below the equal split but well above the self-interest equilibrium. Bob Aumann said, "Al, you must be rich! I'd offer $500,000." Aumann was unwilling to bet so much money on the rationality and self-interest of a random person he was paired with.

goods games accurately): Half are selfish (maximize own earnings $\pi_s$), one-third are Leontief or Rawlsian (maximize $\min(\pi_s, \pi_o)$, where $\pi_o$ is earnings of others), and the rest are utilitarian (maximize $\pi_s + \pi_o$). The representative subject gives as much money to the Recipient as she keeps if the relative payoff for others is about three to four times as large as for herself.

### *2.2.3 Anonymity and Experimenter "Blindness"*

Psychologists have long known that details of experimental protocol or instructions could be taken by subjects as implicit "demands" about what the experimenter intends or hopes to happen. A problem arises if subjects derive utility from "helping" the experimenter by satisfying the demands they perceive. Concerned about these effects, Hoffman et al. (1994) took extreme care to reassure each subject that the experimenter would not know how they behaved in two "double-blind" dictator game experiments.

Dictators got an opaque envelope containing ten dollar-sized blank slips of paper and ten $1 bills. Dictators went, one at a time, inside a large cardboard "phone booth" and took out a combination of ten dollars and blank slips, leaving ten pieces of paper (dollars or slips) in the envelope. (As the dictator left the phone booth the experimenter could not tell from the thickness of the envelope how many dollars were left inside because, even if the subject took all the dollars out, the blank slips were left and simulated the heft of dollar bills.) Then they put the envelope in a large box. After all the envelopes were placed in the box, the experimenter checked all the envelopes to learn the *distribution* of allocations; but since the envelopes were unmarked she did not know what any individual subject had left. Then each Recipient subject took an envelope out of the box and kept whatever money was left.

As Table 2.4 reports, more than half the subjects left nothing and the mean allocation was only 10 percent, significantly less than in control conditions without double-blindness. Hoffman, McCabe, and Smith (1998) also reduced dictator allocations using subtle treatments that they interpret as increasing the 'social distance' between subjects (through instruction changes) or between subjects and the experimenter (through double-blindness) (Frey and Bohnet, 1997). Bolton, Katok, and Zwick (1998) also studied anonymity in dictator games. In their treatment "1 Game 6 Card" in Table 2.4, half the Dictators left nothing and a sixth split the money equally, but they did *not* offer less in the "anonymity" condition.

Bolton and Zwick (1995) imposed experimenter-blindness in ultimatum games. Guaranteeing Proposers that the experimenter will not know their decision is much more difficult in these games, because Proposers need to convey their decision to a specific Responder, *and* find out what

that Responder did; and the experimenter needs to record all these decisions. Their "zero-knowledge" design uses an ingenious scheme, passing boxes back and forth. Tables 2.2 and 2.3 show the results.[7] Anonymity lowers rejections very slightly.

Concerns about experimental "blindness" create both challenge and opportunity. Distancing the experimenter from the subject might undermine experimental credibility, which creates a challenge. Frohlich and Oppenheimer (1997) found that dictator subjects who allocated less were more likely to doubt that there were actually human Recipients on the receiving end (see Table 2.4). Opportunity comes from the fact that anonymity can be easily created in field experiments which complement lab experiments. An example is the "lost letter paradigm" used in the 1950s. Researchers dropped sealed letters containing slugs the size and weight of coins around a town. (Others used "lost wallets" which contain identification so the "recipient" identity can be experimentally manipulated.) The number of letters returned to the experimenter unopened is a measure of altruism. These paradigms sacrifice knowledge of who the subjects are but create experimenter–subject anonymity and have a lifelike feel.

**Summary:** Studies of methodological variables have tested the robustness of ultimatum rejections to the "usual suspect" explanations that are always raised whenever data conflict with economic theory: repetition, stakes, and—a newer concern—anonymity of subjects from experimenter scrutiny. Repetition makes little difference; there is a weak effect of stakes on the rejection of fixed-percentage offers (although subjects reject larger *dollar* offers when stakes go up); and anonymity sometimes lowers Dictator allocations but has little effect in ultimatums.

## 2.3 Demographic Variables

Many researchers are fascinated by the possibility of demographic differences in strategic behavior and social preferences. Gender, academic major, and culture are the demographic variables studied most often, but there is also limited evidence on other variables.

[7] Their results are reported in an unorthodox way to fit in the table. Because subjects do not have a free choice of offer percentages, I used the fraction of times when Proposers offered $x$ instead of the even split $2 to construct a cumulative distribution function (cdf), then used differences in the cdf as estimates of the percentage of subjects making offers in each interval. For example, in the zero-knowledge condition, fifteen of twenty subjects offered the ($3,$1) split instead of an even split, and ten of twenty offered the ($3.40,$0.60) split instead of an even split. This implies that five of twenty subjects (25 percent) would choose an ideal offer, if allowed, between $0.61 and $0.99. Since this range of offers has an average equal to 20 percent of the stake, 25 percent of the offers are reported in the interval [11,20].

### 2.3.1 Gender

It is widely thought that women are more likely to sacrifice their own interests for the sake of preserving harmony in relationships, whereas men are more competitive and apply moral principles that override personal relations (e.g., Gilligan, 1982). The contrast can be seen when children play. If a scraped knee interrupts a game, girls gather around the injured player, sympathizing. Boys are more likely to help the player off the field and keep playing. An economic reason to study gender is the gender gap in wages: Women seem to be paid less for equivalent work (even adjusting for such variables as age, job seniority, and education and skill requirements). Perhaps this gap is partly caused by the different bargaining strategies of women and men, which can be measured in simple experiments (even in the field; see Ayres and Siegelman, 1995).

Eckel and Grossman (2001) measured gender differences in ultimatum bargaining (see also Eckel and Grossman, in press). Some results are summarized in Tables 2.2 and 2.3. Although male and female offers are similar, Eckel and Grossman (and Rapoport and Sundali, 1996) found that women reject less often. Bolton, Katok, and Zwick (1998) and Frey and Bohnet (1995) found no gender difference in dictator games. Solnick (2001) found that both genders demanded more from women, and offered more to men.

Eckel and Grossman (1996b) studied gender in a dictator game with an opportunity for "third-party" punishment. Subjects could divide \$12 evenly between themselves and a type A player who had behaved unfairly toward somebody else in an earlier dictator game, or they could divide a smaller sum $x$ (either \$10 or \$8) evenly between themselves and a fair type B. Females punished more overall and are "better shoppers" (more price sensitive): They punished more often than males when it cost less ($x = \$10$) and punished less often when it cost more ($x = \$8$).

Andreoni and Vesterlund (2001) studied gender in the Andreoni–Miller dictator game with varying values of tokens to a Dictator and Recipient. Overall, women and men allocate the same *dollar* amount to others but this aggregate result hides a big difference: Half the men were purely self-interested, whereas more than half the women were Rawlsian. The mixed effects suggest gender does not have a simple "main effect" on social preferences (such as "women are nicer"—note that they punish more often in the Eckel–Grossman third-party game[8]). Instead, gender seems to interact with many other variables (prices, perhaps beliefs about others), which makes it both a slippery and a rich topic.

[8] Hell hath no fury like a woman scorned?

### *2.3.2 Race*

Modern social scientists are often afraid to study race but it is an interesting variable. Simple games could be used to measure discrimination and whether racial differences account for wage and employment gaps in the economy. Evolutionary psychologists believe that ethnolinguistic differences (which are usually easy to see and hear) are "essentialist" distinctions people are adapted to notice and respond to.

Three studies show interesting distinctions; daring, thoughtful researchers should look for more. In Eckel and Grossman's (2001) ultimatum games designed to study gender effects, they actually found a stronger effect of race: Black students offer more and reject more often. Glaeser et al. (2000) found a small racial effect in their trust games: White students did not repay the trust of Asians. Fershtmann and Gneezy (2001) found a strong difference between behavior toward eastern (Ashkenazic) and western (Sephardic) Jews in Israel; the Ashkenazics were treated more poorly by everybody.

### *2.3.3 Academic Major*

A few studies measure whether students' academic backgrounds affect their allocations. Carter and Irons (1991) ran ultimatum experiments with students majoring in economics and in other fields. Economics majors offered 7 percent less and demanded 7 percent more than others. Because the offer gap did not change when contrasting first-year students and seniors they conclude that the economics-major effect is "born, not made": Students who self-select to study economics tend to behave more self-interestedly (and expect others to), but do not change after four years of economics courses. Other studies have shown mixed effects. Economics and business students offer *more* in Kahneman, Knetsch, and Thaler (1986) and Frey and Bohnet (1995) and behave the same as other majors in Eckel and Grossman (1996) and Kagel, Kim, and Moser (1996).

### *2.3.4 Age*

Developmental studies of how children and adults behave at various ages are important for figuring out whether fairness tastes are innate (as evolutionary theories suggest) or learned through socialization. There are only two studies of age effects. Damon (1980) suggests that children pass through three phases. Before age 5, they are primarily self-interested. From ages 5–7 they focus on strict equality as a way of preventing conflict (even asking to divide a single M&M candy in half to achieve perfect equality when the

number of candies is odd!). After age 7 they begin to think in terms of equity (e.g., rewards proportional to inputs, perhaps coinciding with an increase in cognitive ability to grasp fractions).

To look for these phase changes, Murnighan and Saxon (1998) used children in kindergarten, 3rd grade, and 6th grade. Because of the subjects' young age, fears about comprehension loom large (like cross-cultural research, crossing adult and child cultures). The children divided M&M candies and money. In imperfect information conditions, children did not know how much the other child was dividing; in perfect information conditions they did. The children were not paid, on the advice of teachers. When dividing money, the 3rd graders offered less than the 6th graders (30 percent versus 50 percent). Responders stated an average MAO of 10 percent in the complete information condition, about half as large as in other studies with adults. There were no effects of age on candy offers or responses. However, kindergartners accepted 70 percent of the offers of one penny or one candy, compared with 30–60 percent for the older children.

Harbaugh, Krause, and Liday (2000) did a similar study with 2nd, 4th–5th, and 9th graders in Oregon. Children bargained over ten tokens, which were worth $0.25 each to the 9th graders and could be used by the younger children to buy supplies or toys. Each child played both roles in ultimatum and dictator games. Dictator allocations look like those of adults: About two-thirds of the children gave nothing and the rest gave half or less. Ultimatum offers were the lowest among the youngest children (2nd graders) and got slightly more generous for older children (means were 35 percent, 41 percent, 44 percent). The 2nd graders also were more inclined to accept low offers. There is also a striking effect of *height*: Taller children (adjusting for gender) offered less in the dictator game and the ultimatum games. (There is also a "height premium" in wages and other domains. For example, American Presidents tend to be taller than average; see Persico, Postlewaite, and Silverman, 2001.)

Harbaugh et al. draw a nicely worded conclusion:

> This result gives a new twist to work by others on cross-cultural differences in economic behavior. Explanations of these cultural differences are either really about genetic differences, or they require that there be some way that different cultures persuade people with the same genes to behave differently. We suggest that this process happens in childhood, and we provide evidence of substantial behavioral changes in a sample of children from the same culture, over ages 7 to 14. (2000, p. 20)

Note that in these studies, the youngest children in both studies are *closer* to the self-interest prediction of game theory than virtually any adult population! This is a huge hint that experience does *not* teach people

to behave like payoff-maximizing game-theorists, as is often presumed. If anything, the opposite seems to be true: Grown-up fair-mindedness is the result of the swing of a pendulum from pure self-interest (at young ages), to obsession with strict equality (in 3rd grade), to an adult compromise. These facts cast doubt on a strong version of the hypothesis that an instinct for acting tough in repeated interactions evolved because it was adaptive in our ancestral past. At best, people may have adapted the ability to *learn* to react to perceived unfairness over time (much as a piece of exposed film gradually becomes a picture in a chemical bath; but what chemicals are used affects the exposure). But the innate learning hypothesis is difficult to distinguish from learning that is not innate at all.

### *2.3.5 Brains, Biology, and Beauty*

If ultimatum rejections are mistakes, then subjects who make mistakes in judgment problems should be more likely to reject offers. Clark (1997) tested this hypothesis by having subjects engage in two judgment tasks (probability matching tasks and the Wason four-card logic problem) and a dictator-like allocation task. Subjects who are generous in the allocation task are slightly *better* in the judgment tasks, contrary to the mistake hypothesis. However, reasoning does seem to matter somewhat because Carter and Irons (1991) found that subjects who figured out the perfect equilibrium correctly in the ultimatum game offered and demanded about 5 percent less than others.

Ideally, social preference theories should say something about where preferences come from. An unusual step in this direction is Burnham (1999), who measured testosterone levels ($T$) using saliva samples. $T$ is positively correlated with willingness to behave aggressively, social status, and profession (actors, National Football League players, and firemen are high in $T$; doctors, salesmen, and ministers are low; professors are in the middle, along with the unemployed). Burnham hypothesized that higher-$T$ males have a stronger incentive to preserve their reputation by behaving aggressively in ultimatum games and rejecting offers. In his constrained ultimatum game, Proposers offer either \$5 or \$25 out of \$40. High-$T$s were more likely to reject the \$5 offer, as hypothesized, but were also more likely to choose the generous offer of \$25, contrary to intuition.

Physical attractiveness is economically interesting because there is a well-established "beauty premium" in wages (Hamermesh and Biddle, 1994). To investigate the effect of beauty and gender, Schweitzer and Solnick (1999) had seventy University of Miami students make ultimatum offers and state MAOs. Each subject's picture was taken and rated on attractiveness. Then

the most and least attractive 10 percent of the pictures were shown to a second group, who played ultimatum games against the person whose picture they saw (using the pictured subjects' earlier offers and MAOs).

In the first stage, there was no substantial difference in the offers and MAOs of the most and least attractive subjects. The second group of subjects had a small tendency to offer more to the more attractive subjects, and also to demand more from them. The largest effect is surprising. Men were not especially generous toward attractive women, but women offered about 5 percent more to attractive men than to unattractive men. In fact, the average female offer to good-looking guys was $5.07; this is the only Western group ever found in which the average offer is *more* than half! This extra-fair average results because few women offer less than half to cute guys, and 5 percent offer almost the whole pie ($8–10). The results imply a 10 percent beauty premium (the increase in expected earnings for the attractive compared with the unattractive) and a 15 percent gender premium (men earn more). The fact that the beauty and gender premiums in earnings evident in field data can be reproduced in laboratory bargaining is really interesting and deserves more exploration.

**Summary:** Demographic variables generally have weak effects on ultimatum and dictator behavior, although they are often significant and always intriguing. There are mixed effects of race; very mixed effects of gender and subject academic background (men and economics majors are often more self-interested); and mild effects of testosterone (high-$T$ males reject more often, but are also more generous) and beauty (many women give more than half to attractive guys). The effect of age is strong—young children are more self-interested, then become fair-minded as they grow older. This developmental effect is crucial because it suggests fairness norms are not innate; they change as children develop.

## 2.4 Culture

Culture is a very interesting variable, but cross-cultural comparison raises at least four difficult methodological problems: Stakes, language, experimenter interactions, and confounds.

- *Stakes.* Controlling for stakes requires the experimenter to match the purchasing power of the stake in two different cultures.[9] Converting

[9] In less-developed cultures researchers may have to construct a cost-of-living index by measuring local prices of commonly used items such as pots, knives, radios, sugar, salt, and cooking fat.

a baseline sum into local exchange rates is a good approximation. An alternative solution is to control for stakes in terms of labor supply by equalizing the number of hours of work required to earn the stake amount (Beard, Beil, and Mataga, 2001).

- *Language.* Keeping the meaning of instructions as constant as possible is important. The standard method is "back translation": Have instructions in language A translated to language B, then have the language B version translated back to language A *by a different translator*. If the versions differ, fiddle with the language until it translates back and forth unambiguously.
- *Experimenter effects.* The identity and behavior of the experimenter can sometimes affect what subjects do. Reading from a common script, with as few deviations as possible, controls much of behavior. The biggest mistake in controlling for identity is to use a different experimenter in each culture; then you cannot statistically distinguish the effect of the experimenter from the effect of the culture. The "main effect" of experimenter identity can be controlled by having *each* experimenter conduct an experiment in one culture (e.g., each member of the team of Roth et al., 1991, conducted one session in Pittsburgh), but more care is needed if there are potential experimenter-place interactions.[10] The ideal experimenters speak both languages and are perceived similarly in both cultures (e.g., the first author and experimenter in the Buchan, Johnson, and Croson, 1997, team, who compared Japan and the United States, is Japanese-American).
- *Confounds with culture.* It is extremely difficult to avoid the effects of potentially important variables that are confounded with culture (causing "identification problems" in econometrics terms). For example, suppose you go to two universities in different countries, recruit students from economics classes, and find a difference in their behavior. The difference may be due to culture, or it may be that overall behavior is the same (if people are sampled randomly) but students in one culture are less representative of the population than students in the other culture (e.g., they are older because of pre-college army service, or only the wealthiest go to college). That is, student status may be proxying for unobserved variables that cause the observed behavior, leading to the spurious conclusion that culture matters. The best solution is to match the two cultural samples on as many demographic variables as possible, and measure any variables you can't control (see Botelho et al., 2002).

[10] For example, a female experimenter may be less credible in one culture than another; or there may be a stronger identification between subjects and a same-culture experimenter in one culture than another. The only conclusive control is to have each experimenter do an experiment in each location.

Roth et al. (1991) ran the first thoughtful comparison of bargaining games in America, Israel, Japan, and Yugoslovia. From an anthropologists' point of view, the cultures in these countries are actually very similar, but the paper of Roth et al. still marked an important start. Tables 2.2 and 2.3 summarize results from rounds 1 and 10 in each of the four countries. Offers in the United States and Slovenia were initially more generous and closer together than in Japan and Israel, which were 10 percent lower. By round 10, all offer distributions were more tightly clustered around the initial mean and the 10 percent gap persisted. Players also rejected offers less often in Japan and Israel, especially compared with Slovenia. A key point is that, whereas the Japanese offers are lower and the Israeli offers even lower, rejection rates are no higher in these countries. Roth et al. conclude that "what varies between subject pools is not a property like aggressiveness or toughness, but rather the perception of what constitutes a reasonable offer under the circumstances" (1991, p. 1092).

Buchan, Johnson, and Croson (1997) also compared ultimatum bargaining in Japan and America. They conjectured that the relatively collectivist culture in Japan would exhibit a stronger sharing norm; they were right. Their finding that offers were higher in Japan is the opposite of what Roth et al. observed. The difference illustrates how subtle and interactive cultural effects may be. Buchan et al. used the MAO method and Roth et al. used the specific-offer method, which may have caused differences. A different mixture of students in Japan in the two experiments is another possible explanation.

The most dramatic cross-culture bargaining experiments so far are a remarkable interdisciplinary collaboration between eleven anthropologists and several economists (Henrich et al., 2001, 2002). It began when an enterprising graduate student, Joe Henrich (2000), ran ultimatum experiments during fieldwork with Machiguenga farmers in Peru. He found that Machiguenga offered much less than had been observed in any other subject pool—an average of 26 percent with a mode at 15 percent—and accepted all but one offer! When Henrich came back to UCLA and showed the data to Rob Boyd and me, he wasn't sure whether he had screwed up the experiment or had discovered the first group of people who behave close to the game-theoretic (self-interest) prediction. A close replication in the United States (controlling for stakes) showed the usual result—a huge spike of 50 percent offers—and confirmed that Henrich had found a huge effect of culture.

An anthropology colleague of Henrich's noted that the Machiguenga are quite socially disconnected. The economic unit is the family; families hunt, gather, and practice swidden ("slash and burn") manioc farming. Anonymous transactions within a village are rare. Their society is the opposite of the bar on the television series *Cheers* ("where everybody knows


***Figure 2.2.*** *Map of filed sites for "experimental economics in the bush" project. Source: Henrich et al. (2002).*

your name")—they don't have proper names for other Machiguenga except for relatives. Perhaps the extreme social and economic isolation of the Machiguenga explains why they have no sharing norm. The best way to test such hypotheses, of course, is to compare many cultures. So Henrich and his advisor Rob Boyd assembled a team of anthropologists and one economist to run ultimatum and public goods games in many different cultures which vary in important ways. Figure 2.2 shows a map of the amazing places they did experiments. Figure 2.3 pictures anthropologist David Tracer explaining the ultimatum game to an experimental subject in Papua New Guinea (and her infant who, contrary to appearances, is not for sale).

Their results are described in Henrich et al. (2002). Table 2.5 summarizes some features of the ultimatum game results. In about ten of the cultures (the top rows of the table), average and modal offers are lower than we have seen in many developed countries. Rejection rates are generally low, but vary across cultures. A careful statistical analysis shows that offers are persistently above the utility-maximizing offer (controlling for risk-aversion). Proposers appear to be "rejection-averse" beyond the loss of utility from being rejected. Many subjects said they offered a lot because rejections would cause turmoil in the village. The attentive reader will notice two unusual cultures at the bottom of the table—the Ache headhunters of Paraguay and the Lamelara whalers of Indonesia—who offer *more* than half on average! The anthropologists think these hyperfair offers represent either a norm of

***Figure 2.3.*** *Anthropologist David Tracer explaining the ultimatum game to an experimental subject in Papua New Guinea and her child. Photograph courtesy of David Tracer.*

oversharing because game caught in a hunt cannot be consumed privately, or a potlatch or competitive gift-giving. Accepting an unusually generous gift (such as excess meat caught in a successful hunt) incurs an obligation to repay even more, and is considered something of an insult (since it implies that the giver is a better hunter than the receiver). Hyperfair offers are often rejected, consistent with the competitive gift-giving interpretation. These offers are a reminder not only that self-interest is typically violated in these games, but also that offers and rejections are a language with nuance and cultural variation.

The big payoff from a cross-culture comparison is finding variables that can explain cultural variation. Differences in subject comprehension (rated by experimenters), arithmetic skills, education, anonymity of exchanges, and privacy (i.e., how much the neighbors know about your business) don't seem to matter. Two variables *do* predict differences in offers with a multiple $R^2 = 0.68$: the amount of cooperative activity or economies of scale in production (e.g., collective hunting for whales and big game); and the degree of "market integration." Market integration is an index combining the existence of a national language (rather than a local dialect), the existence of a labor market for cash wages, and the farming of crops for cash. Cultures

***Table 2.5.*** *Summary of ultimatum bargaining games*

| Group | Country | *N* | Stake size | Mean | Mode (percent of sample) | Standard deviation | Rejection rate (percent) | Rejection rate of <20 percent |
|---|---|---|---|---|---|---|---|---|
| Machiguenga | Peru | 21 | 2.3 | 0.26 | 0.15/0.25 (72 percent) | 0.14 | 4.8 | 10 (1/10) |
| Hadza(small) | Tanzania | 29 | 1.0 | 0.27 | 0.20 (38 percent) | 0.15 | 28 | 31 (5/16) |
| Tsimané | Bolivia | 70 | 1.2 | 0.37 | 0.50/0.30/0.25 (65 percent) | 0.19 | 0 | 0/5 |
| Quichua | Ecuador | 13 | 1.0 | 0.27 | 0.25 (47 percent) | 0.16 | 015 | 50 (1/2) |
| Torguud | Mongolia | 10 | 8.0 | 0.35 | 0.25 (30 percent) | 0.09 | 5 | 0/1 |
| Khazaks | Mongolia | 10 | 8.0 | 0.36 | 0.25 | 0.09 | | |
| Mapuche | Chile | 30 | 1.0 | 0.34 | 0.50/0.33 (46 percent) | 0.18 | 67 | 20 (2/10) |
| Au | Papua New Guinea | 30 | 1.4 | 0.43 | 0.30 (33 percent) | 0.14 | 27 | 1/1 |
| Gnau | Papua New Guinea | 25 | 1.4 | 0.38 | 0.40 (32 percent) | 0.19 | 40 | 50 (3/6) |
| Hadza (big) | Tanzania | 26 | 1.0 | 0.40 | 0.50 (28 percent) | 0.17 | 19 | 80 (4/5) |
| Sangu (farm) | Tanzania | 20 | 1.0 | 0.41 | 0.50 (35 percent) | 0.12 | 25 | 100 (1/1) |
| Unresettled | Zimbabwe | 31 | 1.0 | 0.41 | 0.50 (56 percent) | 0.14 | 10 | 33 (2/5) |
| Achuar | Ecuador | 16 | 1.0 | 0.42 | 0.50 (36 percent) | 0.20 | 0 | 0/1 |
| Sangu (herd) | Tanzania | 20 | 1.0 | 0.42 | 0.50 (40 percent) | 0.09 | 5 | 1/1 |
| Orma | Kenya | 56 | 1.0 | 0.44 | 0.50 (54 percent) | 0.092 | 4 | 0/0 |
| Pittsburgh | USA | 27 | 0.28 | 0.45 | 0.50 (52 percent) | 0.096 | 22 | 0/1 |
| Resettled | Zimbabwe | 86 | 1.0 | 0.45 | 0.50 (70 percent) | 0.10 | 7 | 57 (4/7) |
| Los Angeles | USA | 15 | 2.3 | 0.48 | 0.50 (93 percent) | 0.065 | 0 | 0/0 |
| Ache | Paraguay | 51 | 1.0 | 0.51 | 0.50/0.40 (75 percent) | 0.15 | 0 | 0/8 |
| Lamelara | Indonesia | 19 | 10.0 | 0.58 | 0.50 (63 percent) | 0.14 | 20 | 37 |

*Note:* If multiple modes are listed, the first is more common and the second less common. Fractions of the total sample at all modes are in parentheses. For the Lamelara, cigarettes were used (they are like currency) and lower "sham" offers were used to test whether subjects would reject low offers. Reported rejection rates are for the sham offers.

with more cooperative activity and market integration have sharing norms closer to equal splits.

There are important lessons for social science in this project. One is that interdisciplinary research is hard work but worthwhile. The project came together only after Boyd, Henrich, and other anthropologists learned enough about game theory and experimental methods to produce clean data. The anthropologists repay the debt by producing surprises and broadening economists' vision.

The effect of market integration is extremely important. A presumption in economic theory is that, since market exchange can lead to efficient outcomes even if agents are purely self-interested, active markets and self-interest may somehow go hand-in-hand. This project suggests this view might be fundamentally wrong. In cultures with the most market integration, people bargain the least self-interestedly. The anthropologists' very broad view is also a reminder that comparing, say, America and Japan is hardly a study of culture at all because those countries are quite close together in important cultural features.

**Summary:** In their pioneering study, Roth et al. found persistent cross-national differences in ultimatum offers (Japan and Israel are lowest). The key point is that countries have different sharing norms, and comparable rejection rates imply that those different norms are well acccepted (or rapidly learned in the lab) in each country. A remarkable project doing ultimatum (and other) games in a dozen simple societies in remote places such as Papua New Guinea, the Amazon basin, and Africa reveals more dramatic differences across cultures. This project shows some societies in which the self-interested game-theoretic prediction is accurate, and others where there are many "hyperfair" offers that can be interpreted as competitive gift-giving insults. Average offers are strongly correlated with the degree of "market integration," which implies that either market experience creates norms of equal division or the propensity to share evenly permits impersonal markets to flourish.

## 2.5 Descriptive Variables: Labeling and Context

Since Schelling's (1960) work on "psychological prominent" focal points in coordination games (see Chapter 7), it has been well understood that the way in which strategies are described could focus expectations on them and affect the way people play. A related literature in the psychology of decision making shows that the way options are described or "framed" can influence choices. Thus, it is sensible to ask whether alternative descriptions of ultimatum games affect the way they are played.

Hoffman et al. (1994) found that describing an ultimatum game as an exchange—a seller setting a price for a good that a buyer can take or leave—lowers offers by almost 10 percent and leaves rejection rates unchanged. Larrick and Blount (1997) pointed out that ultimatum games are strategically similar to "resource dilemmas" in which players make sequential claims from a fixed common pool of resources and get nothing if their claims add up to more than the pool. When one player makes a claim from the pool, her claim essentially "offers" the second player a chance to claim the remaining amount (or veto it by claiming more), just as in an ultimatum game. They test strategic equivalence by comparing an ultimatum game with a sequential resource dilemma. Offers in the dilemma frame are slightly more generous, and rejections less frequent. They conclude that the language of "claiming" creates a sense of common ownership that makes both sides more generous.

Hoffman, McCabe, and Smith (2000) asked Proposers to "consider what choice you expect the buyer [Responder] to make. Also consider what you think the buyer expects you to choose." These "prompting instructions" raised offers 5–10 percent. They hypothesize that prompting increases Proposer fears of rejection.

**Summary:** Changing the way games are described can have modest effects. Calling it a seller-buyer exchange encourages self-interest. Describing it as a claim from a shared resource pool encourages generosity. There is little doubt that describing games differently can affect behavior; the key step is figuring out what *general* principles (or theory of framing) can be abstracted from labeling effects. Work on framing in matching games, risky choice, and PD games shows how this abstraction might proceed. [11]

## 2.6 Structural Variables

A structural variable changes the way the tree describing the ultimatum game is drawn, typically by adding a move. (Descriptive variables simply change the way the moves or information nodes are labeled.) In my view, structural variables are the most useful to study because they connect simple games to richer economic structures (e.g., adding competition) and also

[11] Work on unpacking focality in matching games (see Chapter 7) points to the roles of distinctiveness and perception. Framing effects in risky choice are understood to occur because of the interaction between shifts in reference points for encoding gains or losses, and systematic gain–loss differences. Pilutla and Chen (1999) found that calling a prisoners' dilemma an "investment game" decreased cooperation compared with a "social event game" description. If players are reciprocal, the labeling could change their beliefs about what others will do and trigger an effect that is self-fulfilling.

provide the most direct clues to the psychology underlying social preference (for example, Konow, 2000, 2001).

### 2.6.1 Identity, Communication, and Entitlement

Some studies varied how much players know about the identity of the person they are playing with, and whether they communicate. Identification may activate empathy or contempt (for example, Jenni and Loewenstein, 1997).

Bohnet and Frey (1999) did experiments using one-way identification in which Recipients held a number in their hands, which paired Dictators in a classroom could use to identify "their" Recipient (but not vice versa).[12] Allowing Dictators to see their Recipient decreased the number leaving zero but did not change the mean significantly. However, when the Recipients stood up and talked briefly about themselves (their name, birthplace, hobbies, and major), the average allocation rose to half and 40 percent of the Dictators give *more* than half. Knowing something about their "charity" seems to activate target-specific[13] sympathy in the Dictators. In a similar study, Eckel and Grossman (1996a) found that allocations doubled when the Recipient was a well-known charity, the Red Cross. In a Swedish experiment,[14] Johannesson and Persson (2000) found no differences between allocations to other students and to members of the general population (other students aren't considered either a good or bad "charity").

Many studies have shown that bargaining face-to-face improves efficiency (see also Roth, 1995b, pp. 295–96). It is not known what components of face-to-face bargaining account for these large effects.

Inspired by earlier research by Hoffman and Spitzer (1982, 1985), Hoffman et al. (1994) allocated the right to propose an offer to the person who answered more general knowledge questions. This shift in "entitlement" lowered offers by about 10 percent (see also List and Cherry, 2000) and reduced Dictator allocations by half. However, it does not appear that this sense of entitlement is entirely shared by ultimatum Responders: rejection rates go up

[12] In earlier experiments (Frey and Bohnet, 1995), both two-way identification between a Dictator and Recipient in a class, and private discussion between the two for ten minutes, doubled mean allocations from 25 percent to about half. However, their design did not control for the possibility of post-experiment interaction.

[13] Frey and Bohnet (1997) studied three-person dictator games in which the Dictator could identify (ID) or talk to one of the Recipients but not the other. When ID and communication were allowed, allocations were about twice as large to the identified Recipient. This means the identification effect is target specific and is not the result of general sympathy toward others activated by looking or talking at one recipient.

[14] The results are interesting from a cross-cultural perspective because Swedes pay high taxes and spend a lot on social services, but their overall rate of Dictator allocation is comparable to that of other countries.

(even in $100 ultimatum games), perhaps owing to self-serving judgments about the legitimacy of entitlement.

### 2.6.2 *Competitive Pressure and Outside Options*

In psychological terms, whether people have behaved unfairly toward us depends on what forces we think caused their unfair behavior (i.e., what "attribution" we make for cause). Careening through a red light is socially acceptable when rushing to the maternity ward, but not when rushing to the video store before it closes, though both acts endanger other drivers equally. The law distinguishes carefully between degrees of accident and deliberation in punishing people for harm to others.

In surveys, consumers say price increases are justified if competition threatens a firm's survival, but not when they exploit a surge in demand (Kahneman, Knetsch, and Thaler, 1986). Using the same intuition, Schotter, Weiss, and Zapater (1996) ask whether competitive pressure could provide an excuse for self-interested behavior. They used two-stage games in which players were allowed to play in a second stage (with a different partner) only if the amount they earned placed them in the top half of earnings among all subjects playing in their role. Dictators certainly used competitive pressure as an excuse—30 percent kept all the money in the two-stage condition, compared with 13 percent in a standard one-stage control. Proposers in two-stage ultimatum games also offered about 10 percent less, and Responders appeared to accept less using the specific-offer method, but not using the MAO method.

Knez and Camerer (1995) added an outside option to ultimatum games (i.e., players earn a nonzero payoff if offers are rejected). If a Proposer's division of a $10 pie was rejected, the Proposer earned $2 and the Responder earned $3. Introducing options creates multiple focal points for how to divide the pie fairly. One focal point is to offer $5, half of the $10 pie. Another is to offer the Responder just enough to get her to accept, such as $3.25 (the self-interested subgame perfect equilibrium). Still another solution is to award each player half the surplus (the gains beyond their outside options), which gives *more* to the Responder ($5.50) if the Responder's option is better.

The rate of disagreement in these games with options is very high—nearly 50 percent, compared with 10–15 percent in most experiments. The high disagreement rate implies that something about the game's structure is undermining the Proposer's willingness to share what the Responder expects, or his ability to guess what Responders will accept. The cause is self-serving bias in judgments of fairness (see Chapter 4): Proposers are more

likely to offer $5 or $3.25, but Responders often state MAO demands for the equal-surplus offer $5.50. However, over five trials the disagreement rate falls somewhat, as Responders' demands fall.

### 2.6.3 Information about the Amount Being Divided

Several experiments have explored how information affects ultimatum offers and rejections. In a typical experiment, Proposers know the exact amount of money to be divided and Responders either know nothing at all or know the probability distribution of possible amounts (see Huck, 1999, for a short review).

Limited information complicates the game in two ways. First, when the pie is unknown the Proposer's offer conveys information to the Responder about the pie size, which complicates the game substantially. Second, if the Responder does not know the Proposer's end of the bargain, the Responder has no way to evaluate whether his share is too low. If Responders accept less when they do not know the Proposer's share, that is very strong evidence that rejections are an expression of preference when they *do* know the Proposer's payoff.

Most studies show that Responders do accept less in the low information condition. Proposers generally do not hesitate to exploit this behavior and offer little when the amount being divided is large. Camerer and Loewenstein (1993) published the first study of this sort. In standard conditions, undergraduate subjects at different schools (Carnegie-Mellon and Penn) made offers and stated MAOs for each of the pies $1, $3, $5, $7, and $9. In the incomplete information condition, Proposers knew the pie size but Responders knew only that the pie was equally likely to be any of the sizes. In both conditions, the median and mean offers were 40–50 percent of each pie. When the pie size was known to Responders, they demanded mean MAOs around 30 percent and the overall disagreement rate was a typical 15 percent. In the incomplete information case, Responders demanded a mean of $1.88 (with substantial variation; many demanded zero). As a result, when pies were small, Proposers could not meet this demand and disagreements were common (39 percent across all pie sizes). Knowing possible pie sizes were $1–9 seemed to focus the Responder's attention on an intermediate pie size, which meant offers were sure to be rejected if the pie was low.

Different results were observed by Mitzkewitz and Nagel (1993), Straub and Murnighan (1995), Croson (1996), and Rapoport, Sundali, and Potter (1996). They all compared bargaining over known pie sizes with incomplete information where players either knew the distribution of possible pies (in Mitzkewitz and Nagel and in Rapoport et al.) or were told nothing at all (in the other studies). In each case Responders seem to give Proposers the

benefit of the doubt: Since a low offer *could* be fair if the pie is small, rejecting it might punish the Proposer unfairly, so lower offers tended to get accepted more often in the face of uncertainty. Proposers generally take advantage of this by offering less. Güth and Hück (1997) ran an interesting study where Proposers knew whether a pie was DM 38 or 16 and a single Responder did not. Responders usually accepted an offer of 8 (half the small pie), but they rejected offers of 7 or 9 half the time, apparently suspecting that an offer that was not *exactly* half of the small pie was probably a small part of the larger pie.

Kagel, Kim, and Moser (1996) ran ultimatum games in which players bargain over a hundred chips with different values to the two players (cf. Roth and Murnighan, 1982; see Chapter 4). Chips were worth $0.10 to one player and $0.30 to another. When only Proposers knew the chip values, they offered around 45 percent when their own chip value was higher, and about 30 percent when the Responder's value was higher, so they are exploiting their information to get more chips for themselves when their value is lower. When the Responders knew the Proposers' chips were more valuable, they tried to squeeze the Proposers to offer more than half in order to equalize dollar earnings; as a result the rejection rate was high (40 percent).

Abbink et al. (2001) ran an ultimatum game in which only the Responders knew how much the Proposer was getting, ex ante, in the event of a rejection. Rejections were more frequent when the Proposer's rejection payoff was lower, which implies that Responders are either envious or deliberately punishing Proposers when it hurts most.[15]

Fairness explanations of ultimatum bargaining leave open the possibility that fairness norms evolve with experience. One way for norms to change and adapt is that players take actions of others as evidence of what is fair or, oppositely, what they can get away with. A way to explore the sensitivity of fair-mindedness to the behavior of others is to provide players with information on what other players have done, and see whether it changes what they do. Two studies show modest effects of social influence. Knez and Camerer (1995) found that ultimatum offers were affected by how much others offered (cf. Harrison and McCabe, 1996b). Cason and Mui (1998) explored social influence in dictator games. Players made two Dictator offers in consecutive rounds. In the second round, players were paired with another subject and told what the other subject offered in the first round (call it $P_1$). There was some social influence, because subjects generally offered more when the other subject's offer was higher.

[15] In Abbink, Sadrieh, and Zamir (1999), rejections are slightly more common when Responders know that the Proposers will know what happened (compared with a condition in which Proposers won't know). But rejections are common even when Proposers won't know they were punished, which means Responders are taking private pleasure in punishment.

### 2.6.4 Multiperson Games

Games with several players raise two important questions: What norms of fair division apply to more than two players? And are players willing to punish unfairness in the same way when their punishment might affect an 'innocent' party? (see Kagel and Wolfe, 2001).

Güth, Hück and Ockenfels (1996) ran a two-stage, three-person game in which the first player learns whether a pie is large (DM 24.60) or small (DM 12.60), then proposes an offer $x$ of that amount to two other players who don't know the pie size. One of the two players can reject $x$ and end the game, or accept it and make an ultimatum offer dividing $x$ with the third player. Notice that sharing the small pie equally among all three means players 1 should offer 8.40 to the other two players. However, when the amount was small only a sixth of the player 1s offer as much as 8.00. When the amount was *large*, 70 percent of the player 1s offer an amount $x$ around 8.40 (pretending the total amount was small). Player 1's exploitation of the other players' uncertainty works because offers are usually accepted. And player 2s generally offered about half of what they were offered to player 3.

Güth and Van Damme (1998) ran a three-person game combining ultimatum and dictator games with and without information about pie sizes. A single Proposer divides 120 points (about $6.80) by making a three-way offer $(x, y, z)$ which gives $y$ to an active Responder and $z$ to an inactive Recipient (leaving $x$ for the Proposer). The Responder can reject the offer, leaving everyone nothing, or accept it on behalf of the inactive Recipient (who does nothing). There are three information conditions: the active Responder knows the entire allocation $(y, z)$ (and can infer $x$), knows only the "essential" information $y$, or knows only the "irrelevant" information $z$. This design tests whether an active Responder cares how much the Proposer allocates to an inactive third party, and what Proposers do in different information conditions.

When the Responder knows her own share, the Proposer offers 30–40 percent of the amount being divided, as in two-person games, and leaves only a little (5–10 percent) for the inactive Recipient. When the Responder does not know her own share, but knows what the inactive Recipient was offered, the Proposer leaves more to the inactive Recipient than in the other cases (12 and 15 percent), trying to signal that she is not too selfish. Overall rejection rates are low (around 5 percent) and Responders don't seem to care much about how inactive third parties are treated.

A dramatic effect of multiple players occurs when there is competition among many Proposers or Responders. Roth et al. (1991) first showed this in their "market games," in which nine Proposers made simultaneous offers to a single Responder (who accepted the highest offer). In the first round offers were dispersed but much higher than in two-person ultimatum games

(most offers are above half). The highest offer was usually large—around 95 percent of the pie—and by the second round virtually all Proposers offered almost all the pie. This result is important because it presents a challenge to theories of the ultimatum game results. As I will note in Section 2.7 below, most theories *do* predict more self-interested behavior in the face of competition. In the market games, a Proposer who is outbid earns no money and ends up being treated unequally. The only way to reduce disadvantageous inequality and earn more is to bid higher. A similar effect of competition is observed when Responders compete (Güth, Marchand, and Rulliere, 1998) and in experimental markets with excess demand or supply (Smith and Williams, 1990; Cason and Williams, 1990).

### *2.6.5 Intentions: Influence of Unchosen Alternatives*

An important effect of attribution of cause (or "intentions") was first shown by Sally Blount-Lyon (1995; then Sally Blount). First note that a Responder faced with an (8,2) split, for example, might reject it for two different reasons: She might dislike being treated unfairly by somebody who benefits from unfair treatment; or she might just dislike unequal payoffs. If she dislikes unfairness but does not mind inequality, then she will reject the (8,2) division if it was made by a Proposer, but accept the same (8,2) division made by a random device or third party (e.g., a court or regulation). Indeed, Blount-Lyon found that players stated lower MAOs when offers were made randomly than when they were made by Proposers.

Falk, Fehr, and Fischbacher (in press) investigated whether "intentions" matter using three discrete ultimatum games (see also Brandts and Sola, 2001; and Andreoni, Brown, and Vesterlund, 2002). A Proposer can offer either (8,2) or one of the divisions (5,5), (2,8), or (10,0). The question is how Responders react to the (8,2) offer depending on which of the unchosen alternatives was possible (see Table 2.6). Each alternative generates a different psychological evaluation of the intentions lying behind the offer (8,2). Offering (8,2) instead of (5,5) is relatively unfair; the Proposer's intentions are selfish. Offering (8,2) instead of (10,0) means the Proposer intends to be nice, helping the Responder get something instead of zero. Offering (8,2) instead of (2,8) means that the Proposer had to choose between two lopsided divisions, and chose the one which was better for her (she did not intend to take the short end of the stick if somebody had to do so).

The relative frequencies of (8,2) offers and rejections of that offer are shown in Table 2.6. Intentions matter because frequencies of rejection of (8,2) do vary substantially with the unchosen offer. When the unchosen offer is the equal split (5,5), nearly half the Responders rejected (8,2). When the unchosen offer was the selfish (10,0), only 10 percent rejected (8,2).

*Table 2.6. Ultimatum games with varying unchosen paths*

| Unchosen offer | Interpretation of (8,2) offer | How often the (8,2) offer is | |
|---|---|---|---|
| | | Rejected | Proposed |
| (5,5) | Relatively unfair | 0.44 | 0.31 |
| (2,8) | Not sacrificial | 0.27 | 0.73 |
| (8,2) | Neutral | 0.18 | — |
| (10,0) | Relatively fair | 0.09 | 1.00 |

*Source:* Falk, Fehr, and Fischbacher (in press).

The Proposers seem to anticipate these differences in likely rejection rates, because they offer (8,2) only a third of the time when the alternative is the equal-split (5,5), and offer (8,2) most of the time in the other conditions. Since forgone alternatives matter, theories that evaluate social preferences by applying a utility function to terminal node outcomes are incomplete (see Section 2.7).

**Summary:** Changing structural variables in simple games is very helpful for understanding what is going on and extrapolating to complex settings. Creating entitlement by letting a contest winner be the Proposer lowers offers. Knowing who you are allocating money to and hearing them talk raises average Dictator allocations, and many give much more than half. (Fund raising agencies know this and create "identifiable victim" sympathy with tragic pictures of victims of disaster and hunger.) When Responders don't know how large the pie being divided is they usually accept less because they are reluctant to reject low offers that might be fair offers from small pies (even if they are probably unfair offers from large pies). Proposers take advantage of this "benefit of the doubt" by offering less.

Multiperson games show that the social preferences are not based on judgments about another player's overall generosity, but are based on judgments about another player's fairness toward oneself. For example, in three-person ultimatum games, active Responders accept offers that give very little to an inactive third party.

When Proposers and Responders compete, there is no way for fair-minded players to earn money *and* enforce fairness, so outcomes consistent with self-interest result (in practice and in theory). This does not imply that players are self-interested in markets per se. It just means that in some exchange institutions it is impossible to express social preferences for fairness or low inequality, so only preferences for earning more money (self-interest) are expressed.

Some studies show that intentions—the influence of unchosen alternatives on the psychological evaluation of a chosen alternative—matter. In ultimatum games with competitive pressure (high-earning Proposers get to play again), the need to compete is an excuse for Proposer greed, so Responders accept lower offers. Responders are also more likely to accept unequal offers generated by a random device than by a Proposer who benefits from the inequality; they are punishing the person who benefits from the unfair action rather than simply rebelling against inequality.

Many of these findings are tentative. They deserve replication and further exploration because they will put a lot of empirical constraint on theories of social preference. Good theories must not just explain why some players reject ultimatum offers. They must also explain why randomly generated offers are accepted more often, why the nature of unchosen offers affects rejections, why competing Proposers give everything to a single Responder, why Responders don't care how much is offered to a silent Responder, and so forth. A theory that can explain a broad range of these phenomena will be extraordinarily useful in designing organizational contracts, explaining intrahousehold bargaining, understanding everyday concepts of fairness and justice, and so forth. These phenomena also suggest a way to put variables from psychological and sociological theories into a game-theoretic language that could unify some aspects of social sciences.

## 2.7 Trust Games

Trust is an elusive concept which cuts across many disciplines. Marriage therapists talk about rebuilding trust after an affair. Sociologists are interested in how social networks produce or inhibit trust. Economists emphasize trust as a way of reducing the costs of transacting, "lubricating" the economy (e.g., Arrow, 1974). Whenever there are gains from trade, there is a productivity advantage to any institution or norm that assures traders that the other side will hold up their end of the deal. Legal contracts, third-party assurance (e.g., a mutual friend vouches for the trustee's honor), family solidarity, and threats of violence can provide assurance but they cost something. Trust is cheap.

Development economists and pundits think differences in trust (or "social capital") can explain why some countries or cultures are rich and others struggle. For example, Knack and Keefer (1997) found a strong cross-country correlation between economic growth and the fraction of citizens who said they generally trust people. Such explanations will live or die on accurate measurement of trust and social capital. But, as Putnam (1995) laments, "Since trust is so central to a theory of social capital, it would be

desirable to have strong behavioral indicators of trends in social trust or misanthropy. I have discovered no such behavioral measures."

Here are some examples of trust:

1. When I was visiting the University of Iowa from Chicago, people there constantly told me how friendly small-town living was compared with life in the big city. Driving to the airport, my host's car had trouble. He pulled over at a nearby farm and peered under the hood helplessly. A farmer came out in his truck to see what the problem was. My host, who didn't know the farmer, explained that he had to get me to the airport quickly. The farmer then *lent us his truck* to drive to the airport. Before we hopped in the truck, my host went to hand over the keys to his sleek, ailing BMW to the farmer as collateral. The farmer said "Hell, just keep 'em, I wouldn't know what to do with those anyway. Just come on back and we'll take care of your car."
2. Tokyo's lost and found center is famous for its ability to return lost items to their owners (*Los Angeles Times*, 1999b). About 72 percent of items (measured by value) are returned to their owners, including a bag with $90,000 in cash. The system mixes a tradition of teaching children to turn in lost property at police kiosks (*koban*), a golden rule that creates empathy for how the owner of the lost item must feel, and straightforward incentives (by law and social convention, the owner must give 5–20 percent of the item's value to the finder as a thank-you).
3. Why do firms prefer to lay off workers during a downturn rather than cut wages? Since risk-averse workers should prefer wage cuts, the preference for layoffs is a major puzzle. Managers inexorably refer to the fact that workers get upset when wages are cut. Since workers have a lot of leeway to do good or bad work, keeping wages up is a kind of insurance to guarantee reasonable work performance (see Bewley, 1998). Similarly, despite the rise of the "war on drugs," during record lows in unemployment in the late 1990s firms were reluctant to test employees for drug use because "drug testing, particularly without probable cause, seems to imply a lack of trust and presumably could backfire if it leads to negative perceptions about the company" (*Los Angeles Times*, 1999a). Furthermore, drug testing is most common in lower-level white- and blue-collar jobs, even though drug use surely harms overall productivity much more in professions such as movie direction, surgery, floor trading on Wall Street, and politics. But firms say mandatory drug testing would upset these high-level professionals and undermine trust.
4. Noncontractual reciprocity occurs even in a domain that is notorious for greed and ruthlessness—Hollywood. When the film *Titanic* was running spectacularly over budget, many critics forecast the biggest flop

in film history. Director James Cameron surrendered part of his fees and "points" (profit percentage) to comfort panicky studio executives (a classic example of signaling; see Chapter 8). The film went on to become the third-highest-grossing film of all time, and the fees and points Cameron waived had lost him $50 million. Giddy studio executives gave Cameron the biggest tip in history by returning those fees. Increases in director fees of this magnitude never occur just because films do well and studios want the director to make more hits for them (as repeated-game models could explain). What was special was the fact that Cameron had waived the fees to begin with. The studio's payment was clearly a repayment of his sacrifice.

A beautiful simple game to measure trust was proposed by Berg, Dickhaut, and McCabe (1995; and earlier by Camerer and Weigelt, 1988). In their game, the Investor has $X$, which she can keep or invest. Suppose she invests $T$ and keeps $X - T$. The investment of $T$ earns a return, at a rate $(1 + r)$, and becomes $(1 + r)T$. Then another player, the Trustee, must decide how to share the new amount $(1 + r)T$ with the Investor. (The Trustee is playing a dictator game in which the amount to be allocated was determined by the Investor.) Suppose she keeps $Y$ and returns $(1 + r)T - Y$. Then the total payoffs are $Y$ for the Trustee and $(X - T) + (1 + r)T - Y$ for the Investor, which is $X - Y + rT$.

In this game, trust is the willingness to bet that another person will reciprocate a risky move (at a cost to themselves). Trust is risky because the Investor will regret having entrusted if she doesn't get much back.[16] $T$ measures the amount of trust. The amount returned, $(1 + r)T - Y$, measures trustworthiness. If players maximize their earnings, the Trustee will keep it all ($Y = (1 + r)T$); the game has moral hazard or hidden action which cannot be guaranteed contractually. Anticipating this, a self-interested Investor should keep the money rather than invest it.

Trust must be risky. Trustworthiness must also go against the Trustee's self-interest, to test whether people are willing to sacrifice to satisfy moral obligation.[17] Sociologists and psychologists usually object that this game doesn't capture all there is to trust because the two-person one-shot game does not include relationships, social sanctions, communication, and so many other rich features that may support or affect trust. That's precisely the point—the game requires *pure* trust. It also provides a plain benchmark

[16] A prisoners' dilemma played sequentially is also like a trust game; the first player cooperates, exposing herself to possibly earning the lowest payoff, trusting the second player to reciprocate by cooperating back.

[17] Otherwise the game does not measure trust, it is simply a bet on the rationality and self-interest of others, like the Beard–Beil games in Chapter 5.

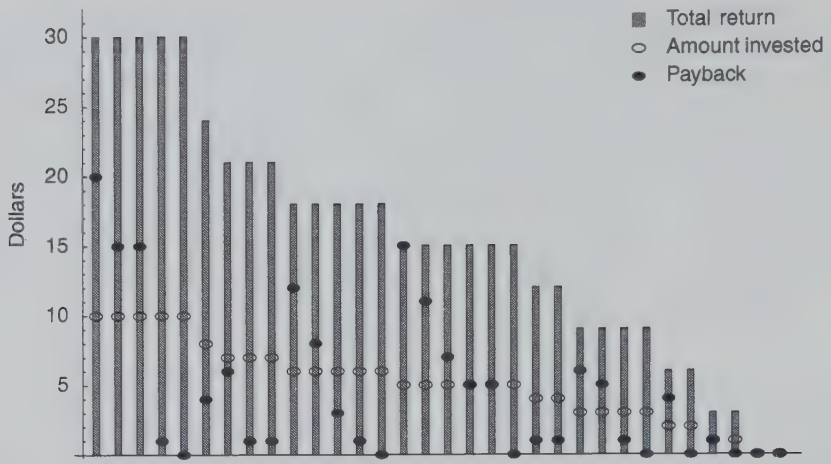



*Figure 2.4.* *Investment and repayment in a trust game. Source: Based on Berg, Dickhaut, and McCabe (1995).*

against which trust under more complicated conditions can be compared, like a wire mannequin that can be dressed up with clothing.

The trust game caught on quickly. Berg, Dickhaut, and McCabe (1995) played it with an initial amount $X = \$10$, a rate of return $r = 2$, and an elegant mailbox design to ensure "double-blindness." Their results are shown in Figure 2.4. Points are arrayed from left to right according to the amount of the investment. The open circle in each bar shows what was invested ($T$). The height of the bar is the amount available to split, $3T$. The dark circle shows how much was returned. If the dark circles are above the open circles, then trust is repaid. On average, Investors put in about 50 percent; five of thirty-two invested it all and only two invested none. The average amount repaid was about 95 percent of what was invested (or, equivalently, about a third of the tripled amount), with a wide dispersion—half repaid either nothing or an insulting \$1. The fact that the return to trust is around zero seems fairly robust (e.g., Bolle, 1995).

The substantial amount of blind trust and trustworthiness Berg et al. observed among students in Minnesota has been generally replicated in various places, with some interesting variation. In Massachusetts, Ortmann, Fitzgerald, and Boeing (2000) experimented with variants of the Berg et al. design, prompting subjects by asking how much they expected and giving "social history" about what others had done. Investments ranged from 40–60 percent across treatments and repayments averaged 110 percent. Koford

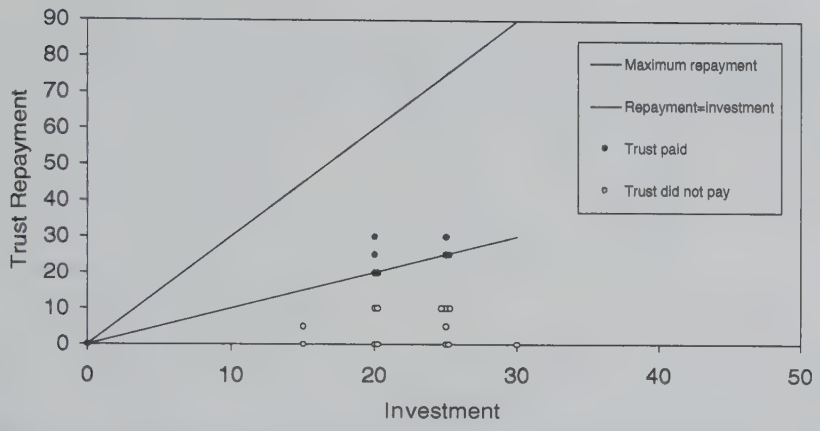



***Figure 2.5.*** *Investment and repayment in a trust game in Kenya. Source: Based on data from Ensminger (2000).*

(1998) found Bulgarian students were surprisingly[18] trusting, investing 70 percent and getting 150 percent back. He speculates that Bulgarians are used to trusting among themselves precisely because their trust in authority is so low. Willinger, Lohmann, and Ususnier (1999) found the French trusted much less that the Germans, but both nationalities returned about 40 percent. Ensminger (2000) found very little trust and trustworthiness among Orma herders in Kenya (see Figure 2.5). In Figure 2.5, closed circles represent repayments greater than investment (trust paid); open circles represent repayments less than investments. The Orma invested 40 percent of 50 currency units—only one invested more than half; and they repaid only 55 percent (so most of the points in Figure 2.5 are open circles close to the zero-repayment $x$-axis). Note that Kenya is considered one of the more corrupt countries in the world, measured by indices of "transparency," which guess the extent of bribery, bureaucratic corruption, and black market trade, so it is encouragingly consistent that this simple game shows low levels of trust also.

Jacobsen and Sadrieh (1996) conducted a trust experiment in Germany in which subjects made decisions as groups of three and their discussions

[18] Subjects were thirty-two undergraduate students in economics and business at Sofia University playing for 1,000 leva (roughly two hours' wages). Koford reports Bulgaria is a typical East European country with little trust in authority and high levels of fraud, corruption, and bribery, including widespread cheating on exams, and professors accept bribes to give good grades.

*Table 2.7. Payoffs in a trust game*

| | Trustee move | |
|---|---|---|
| Investor move | Repay trust | Don't repay trust |
| Don't trust | P, P | P, P |
| Trust | R, R | T, S |

*Source:* Snijders and Kerens (1998).
*Note:* $S < P < R < T$.

were videotaped. They invested 60 percent and were repaid 110 percent. In their discussions, all Investor groups mentioned the chance of earning more by investing, most discussed whether to invest all or none, and about half talked about charity and altruism toward the other side. *All* Trustee groups talked about reciprocity (often using the German verb *honorieren*). The students were well aware that there was risk of moral hazard in a one-shot anonymous game (though some fantasized about possibilities for post-experiment punishment), which casts doubt on the popular hypothesis that they do not distinguish between one-shot and repeated games.

Snijders and Keren (1998) varied payoffs across binary-choice trust games with the structure (in the normal form) in Table 2.7. If trust is not repaid, the Investor earns a payoff $S$ below the no-trust guarantee of $P$ (like the "sucker" payoff in PD) and the Trustee earns the highest payoff, $T$. If trust is repaid, both earn $R$ ($> P$, so reciprocated trust helps both sides). They derive some propositions from a simple "social orientation" model proposed by Edgeworth (1881) and studied extensively by social psychologists in the 1970s, $u_i(x_i, x_j) = x_i + \alpha_i x_j$.[19]

The model predicts trust will be lower when "risk" $(P - S)/(R - S)$ is high and "temptation" $(T - R)/(T - S)$ is high (risk is probably influenced by anticipated temptation, of course), and players will be less likely to repay trust when temptation is high. (The Fehr–Schmidt model of inequality-aversion, see section 2.8.2 below, also points to the temptation factor $(T - R)/(T - S)$ as the key predictor of Trustee behavior.[20])

Across thirty-six payoff structures, a probit regression of the dummy variable "trust" showed a strong effect of risk and a smaller effect of temptation. (Subjects who carried an organ donor card also trusted more.) Regressions

[19] See McClintock (1972). Like simple altruism models, it is incomplete because the social orientation weight $\alpha_i$ can empirically change sign depending on the game and beliefs about a player's opponent. The great triumph of new models such as Rabin's (1993), discussed below, is that they endogenize $\alpha_i$ in a parsimonious, falsifiable way.

[20] In their model, the Trustee repays iff $T - \beta(T - S) < R$, or $\beta < (T - R)/(T - S)$, where $\beta$ is a coefficient measuring the disutility of advantageous inequality (or guilt).

of the dummy "repay trust" showed a strong effect of the temptation ratio $(T - R)/(T - S)$, as predicted by their model and by Fehr–Schmidt, and a weak gender effect (men are less trustworthy). An important fact is that including a dummy variable for each game improves fit very little, which means virtually *all* the impact of the payoff variables is captured by the risk and temptation measures.

Van Huyck, Battalio and Walters (1995, 2001) studied a trust game modeled after a "peasant" who must decide how much to plant (at various possible rates of return $r$) when a Dictator landowner decides how much to confiscate by "taxation." In their discretion condition, Dictators announce a tax rate *after* the peasants' decisions (as in the trust games above). Peasants believe they cannot trust the Dictators, so they invest very little (the median is zero) and the Dictators usually grab everything.[21] In another condition, when Dictators can precommit to tax rates they usually choose rates that induce efficient investment and create gains for everyone. Median investment is 100 percent and median tax rates are close to the efficient level, although Dictators share surplus more evenly than predicted when returns $r$ are high. In a third reputation condition, Dictators have discretion but are matched repeatedly with the same peasant (with a continuation probability of 5/6). Allowing reputation building creates outcomes midway between the discretion and commitment conditions, with some convergence to optimality over time. These results are important because they find the *least* trust seen so far in the discretion treatment, and also show that reputation in repeated matching is a partial substitute for precommitment.

### *2.7.1 Is Trustworthiness Just Altruism?*

Trustee repayments are usually thought to reflect moral obligation or positive reciprocity toward Investors who took a risk that benefits the Trustee. But this conclusion is right only if repayments from $X$ by Trustees are larger than allocations in a game where a dictator allocates a sum $X$ that did not result from an Investor move. Two studies suggest, surprisingly, that only a small amount of the repayment by Trustees is owing to positive reciprocity.

Dufwenberg and Gneezy (2000) studied a trust game in which the Investor could let the Trustee divide 20 Dutch guilders between them, or could take an outside option and earn $x$ for herself, ranging from 4 to 16 Dutch guilders (giving nothing to the Trustee). The average amount Trustees returned is one-third, insignificantly larger than the average allocation of 30 percent in a 20-guilder Dictator game. Furthermore, the amount Trustees

[21] The Dictator's dilemma is sometimes called the "paradox of omnipotence": A Dictator above the law is unable to commit herself to be unable to do something in the future.

repaid did not depend significantly on the sure option value $x$ that the Investor passed up to the Trustee (and herself).

Cox (1999) also compared trust repayments $r(3t)$ (given investments of $t$ determined by Trustees) with the amounts allocated out of $3t$ in dictator games. (That is, the size of the Dictator pie was matched to the amount Trustees repaid in the trust game.) The difference between repayments and allocations, $r(3t) - 3t$, is significantly positive ($1.20) but small (around 10 percent), which indicates only a small extra effect of positive reciprocity. Cox also compared decisions of individuals with their later decisions as part of a three-person group. Groups give and repay less and appear to be disproportionately influenced by the least trusting and trustworthy members.

### *2.7.2 Indirect Reciprocity, Karma, Culture*

In the trust games discussed so far, Investors know that the Trustee they are paired with will repay money to them (or not). Many social transactions are less direct. For example, dealing with bureaucrats, firm employees, or spouses in a household, there is usually a presumption that a repayment that is agreed to by one person will be honored by his or her spouse or fellow bureaucrat or employee if the original Trustee is not around.

To measure the strength of this sort of "indirect reciprocity," two studies see what happens when Investors know they will be repaid money by a different Trustee than the one they entrusted their money to. Buchan, Croson, and Dawes (2000) used three experimental conditions.

1. The *pairs* condition is a standard control replicating the many studies described above.
2. In a *foursome* condition, there were two Investors and two Trustees but each Trustee repaid the opposite Investor (i.e., A invested with B and C invested with D but D repaid A and C repaid B). That is, a person was repaid by the Trustee whom "their" Trustee's Investor invested with (try saying that three times quickly).
3. In a *society* condition, Investors and Trustees were in separate rooms and which trustee paid money to a particular Investor was random.

The group and society conditions model situations in which moral obligations to repay are passed around, as if Investors believe in "karma" (good deeds will eventually be repaid by somebody else doing a good deed). Like blind trust itself, karma sounds silly to economists but many religions include some concept like it; and note that a society in which karma is widely believed in will be more productive than one without it.

Buchan et al. also measure how sharply trust drops off from pairs to foursomes to societies in different countries (United States, Japan, Korea, and China). Their interest in these countries is motivated by claims such as the popular 1980s' (pre-bubble collapse) idea that trust was responsible

*Table 2.8. Trust game results across three conditions and four countries*

| Countries | Pair | Foursome | Society | Overall |
|---|---|---|---|---|
| *Fraction invested* | | | | |
| American–Chinese | 0.76 | 0.49 | 0.49 | 0.54 |
| Japanese–Korean | 0.51 | 0.48 | 0.28 | 0.41 |
| Mean | 0.64 | 0.48 | 0.39 | 0.47 |
| *Fraction of tripled investment returned* | | | | |
| American–Japanese | 0.28 | 0.13 | 0.11 | 0.15 |
| Chinese–Korean | 0.41 | 0.25 | 0.18 | 0.25 |
| Mean | 0.35 | 0.19 | 0.15 | 0.20 |

*Source:* Buchan, Croson, and Dawes (2000).

for fast economic growth in Japan, the idea that the non-market-oriented Chinese might be less trusting, and the importance of family and business groups as economic units in the Asian countries.

Table 2.8 (pooling separately for each measure across countries with similar results) shows that the Chinese were most trusting and trustworthy, and the Japanese least. Overall investment was about 60 percent and the amount repaid was 105 percent of the investment, quite close to other results reported above. Trust and repayment did drop off in the group and society conditions in most countries, although there is substantial belief in karma even when the Trustees didn't know who would repay them.

Dufwenberg et al. (2000) also compared pairs and foursomes (which they call direct and indirect reciprocity) with $r = 2$. In the pair condition, they added incomplete information about whether the return was $r = 1$ or 3 (known to the Trustee but not the Investor), to see whether the Trustees would give less when they could hide behind the possibility of a small return as an excuse (as we observed in ultimatum games). Results are shown in Table 2.9. Trust and repayment are essentially the same in all conditions and comparable to amounts seen in earlier studies.

*Table 2.9. Trust game with pairs and foursomes*

| | Complete information | | Incomplete information |
|---|---|---|---|
| | Pair | Foursome | Pair |
| Fraction invested | 0.60 | 0.53 | 0.55 |
| Fraction returned | 0.28 | 0.37 | 0.26 |

*Source:* Dufwenberg et al. (2000).

*Table 2.10. Contribution rates for high-cost conditions*

| Previous contributions | History condition | | |
|---|---|---|---|
| | Donor | | Recipient history |
| | History | No history | |
| 0 of last 6 | 0.02 | 0.00 | 0.24 |
| 1 of last 6 | 0.42 | 0.21 | 0.45 |
| 2 of last 6 | 0.48 | 0.31 | 0.51 |
| 3 of last 6 | 0.65 | 0.44 | 0.60 |
| 4 of last 6 | 0.71 | 0.58 | 0.73 |
| 5 of last 6 | 0.78 | 0.70 | 0.77 |
| 6 of last 6 | 0.85 | 0.98 | 0.94 |
| Overall | 0.70 | 0.25 | |

*Source:* Seinen and Schram (1999)

Seinen and Schram (1999) study indirect reciprocity, inspired by biological thinking about how an organism's "image"—its observed record of cooperativeness—induces others to cooperate with it (e.g., Nowak and Sigmund, 2000). The idea is similar to repeated-game equilibria in which a player *i* anticipates reciprocity by an unknown future player who conditions her contribution on whether *i* contributed. In each period, a player could donate 250 to another randomly chosen player at a cost of 150 to themselves.[22] In a history condition, players were told the last six decisions made by their target recipient (and knew their own history would be reported). In the no-history condition, no such history was reported.

Table 2.10 shows that reporting history has a huge effect, raising the contribution rate from 25 percent to 70 percent (Wedekind and Milinski, 2000, replicated this result). The middle two columns show that contribution rates by a donor are strongly increasing in that donor's own history (this reflects persistent individual differences; some just donate a lot and others do not). The right column shows that donors give more often to those recipients who contributed recently. There is also a small decay in contributions over time (as in most public goods games with stranger matching). Reliable individual differences can be used to forecast later contributions with 75 percent accuracy.

[22] Their game is a dictator game with a multiplier greater than 1 on allocations made to others. They also studied a low-cost condition in which the donor's cost was 50. Overall contribution was higher (86 percent than in the high-cost condition but the image results are basically the same so the text reports only the high-cost results.

### 2.7.3 A Complex Omnibus Game

McCabe, Rassenti, and Smith (1998) used a complex Big Tree game which illustrates concepts of reciprocity. Many of the experiments in this book (and particularly this chapter) use a decomposition principle in design: Pick a game that isolates which features of game theory are violated if predictions prove wrong.[23] The Big Tree game illustrates an opposite design strategy: Use a complex game where several forces operate at once. The top panel of Table 2.11 sketches the payoffs.[24] Player 1 either chooses an outside option which ends the game (it is a nuisance and rarely picked) or gives the move to player 2, who chooses a left or right branch (the left and right halves of Table 2.11). Choosing right gives player 1 a chance to take a (30,60) node or pass to player 2, who can choose (40,40) or pass back to player 1, who has two bad choices ((15,30) and (0,0), which is omitted from the table). Choosing left gives player 1 a chance to enforce (50,50) or pass to 2, who can choose (60,30) or pass back to player 1 for a poor choice of (20,20) and (0,0) (omitted). Assuming self-interest, the subgame perfect prediction (in italics) is that player 1 rejects the option, player 2 moves right, and player 1 passes to 2, who chooses (40,40). Note, however, that players can earn (50,50) if player 2 moves left and player 1 chooses the (50,50) endnode (although 1 will pass if she believes 2 will take the self-interested outcome of (60,30)). This reciprocity-based outcome is marked in bold. A version of this game called game 2 flips around the location of the (50,50) and (60,30) outcomes on the left branch of the tree so that (60,30) comes first and can be chosen by player 1 right away.

Their game pits the self-interested subgame perfect outcome (right, (40,40)) against a loose concept of reciprocity which leads to the (left, (50,50)) outcome. The paper is informal about game-theoretic details. The authors compare "one-shot" games, with a partner-matching treatment (called "same"). In a treatment where players know only their own payoffs ("private payoff information"), social preferences and reciprocity are disabled.

The overall frequencies of player 2 choosing left or right, and conditional frequencies of each of the other possible outcomes, are shown in Table 2.11. The reciprocal and subgame perfect outcomes are reached about equally often in one-shot games and reciprocity is much more

[23] For example, the PD game is a *bad* decomposition because defectors could be self-interested, envious, or conditionally cooperative and pessimistic; so defection, by itself, does not pin down a defector's motives. The ultimatum is a *good* decomposition because rejections by Responders isolate negative reciprocity (if they reject more often when offers come from Proposers rather than random devices). Rejections are a clear indication of negative reciprocity and nothing else (e.g., they have nothing to do with failures of strategic thinking, as in Chapter 5, or reputational considerations, as in Chapter 8).

[24] Terminal nodes that result in (0,0) are omitted because they are rarely chosen.

***Table 2.11.*** *Conditional frequencies of choices in multimove game*

| | Payoffs (player 1 on top, player 2 on bottom) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Left branch | | | | Right branch | | | |
| Player move | | 1 | 2 | 1 | | 1 | 2 | 1 |
| | | 50 | 60 | 20 | | 30 | 40 | 15 |
| Treatment | % left | 50 | 30 | 20 | % right | 60 | 40 | 30 |
| One-shot 1 | 43 | **0.85** | 0.08 | 0.08 | 57 | — | *0.94* | 0.06 |
| One-shot 2 | 46 | **0.50** | 0.50 | — | 54 | 0.00 | *1.00* | — |
| Same 1 | 82 | **0.88** | 0.04 | 0.06 | 18 | 0.04 | *0.85* | 0.05 |
| Same 2 | 62 | **0.84** | 0.16 | — | 38 | 0.17 | *0.70* | 0.13 |
| Private payoff information 1 | 16 | **0.43** | 0.56 | 0.02 | 84 | 0.14 | *0.84* | 0.02 |

*Source:* McCabe, Rassenti, and Smith (1998).
*Note:* Subgame perfect outcomes are in *italics*; reciprocal outcomes are in **boldface**.

common in partner-matching. When payoff information is private, however, reciprocity falls apart and the subgame perfect (40,40) outcome is usually reached.

### *2.7.4 Multistage Trust Games*

In multistage games, trust can usually be supported by repeated-game reputation building, so multistage games are not evidence of blind trust and therefore make a different point than the studies in this chapter do. However, some experiments on multistage trust games are described in Chapters 5 and 8, and two new studies of "centipede" games are worth noting here.

A centipede game is a multistage trust game in which players alternate moves. Each player can "take," ending the game, and take some percentage of the available surplus. Or the player can "pass," increasing the surplus, and allow the other player a chance to take or pass. The game ends with a terminal node where one player must take. Because there is a terminal node, self-interested players will always take at the end. If players backward induct, this leads to unraveling—taking at *all* nodes. The typical finding (McKelvey and Palfrey, 1992; see Chapter 5) is that players pass until a couple of steps from the end. (Play in finitely repeated prisoners' dilemmas, a cousin of centipede games, is similar.)

Ho and Weigelt (2000) did four-move centipede games in which the surplus doubled with each move, players could take as much as they wanted at each node, and the terminal node is (0,0). Subjects gave normal-form strategies (i.e., they stated in advance whether they would take at each

possible node). They did experiments in China, America, and Singapore (which is quite Westernized, a convex combination of China and America). They observed much more self-interest than in other studies—about 30 percent (50 percent) of player 1s (2s) took at their first node, and players took almost all the surplus (95 percent).

Rapoport et al. (in press) ran a three-person centipede game with nine nodes for very large stakes (subjects could have earned $1,500 if everybody always cooperated). Like Ho and Weigelt, they observed early taking consistent with self-interest—one-third of the games ended at the first or second node. These two studies are inconsistent with what McKelvey and Palfrey found (and see also Chapter 8 results); the difference is probably due to the presence of a (0,0) terminal node.[25] In the study by Rapoport et al., playing with *two* other players, rather than one, is a big psychological step, because it requires a player to trust each of two others, *and* to trust that the others both trust each other, and so on.

### 2.7.5 *Gift Exchange in Experimental Labor Markets*

Fehr and many collaborators did a series of experiments on labor markets (summarized in Fehr and Gächter, 2000b). Their studies illustrate the virtues of a steady, ambitious, cumulative research program. In their paradigm (e.g., Fehr, Kirchsteiger, and Reidl, 1993) firms offer a wage $w$ and workers who accept it then chose an effort level $e$. Firms earn $(q - w)e$ and workers earn $w - c(e)$, where $c(e)$ is a convex cost function over effort levels from 0.1 to 1.0. There is an excess supply of labor (eight workers and six firms). In most of their experiments, firms post offers which workers accept or reject (usually in random order). Firms cannot contract with specific workers and do not know their identities, so workers cannot build up reputations. The experiments have ten to twenty periods, to observe equilibration over time.

Once workers are hired, firms cannot control workers' effort and self-interested workers will choose the minimal effort (0.1). Anticipating this, profit-maximizing firms should pay the market-clearing wage of 1. An alternative theory is "gift exchange" (Akerlof, 1982): Workers reciprocate a firm's "gift" of an above-market wage by exerting more effort than they have to. This reciprocation is *not* a repeated-game equilibrium because workers

[25] Many explanations of passing in centipede games rely on the idea that some fraction of "nice" players will pass even at the last node, owing to social preferences, and self-interested players have a reputational incentive to mimic the nice ones until near the end (Camerer and Weigelt, 1988; McKelvey and Palfrey, 1992). But there is no sensible social preference that would explain passing from an advantageous division for oneself to a (0,0) payoff. Without the possibility of such niceness, self-interested players have no incentive to build trust in early stages.

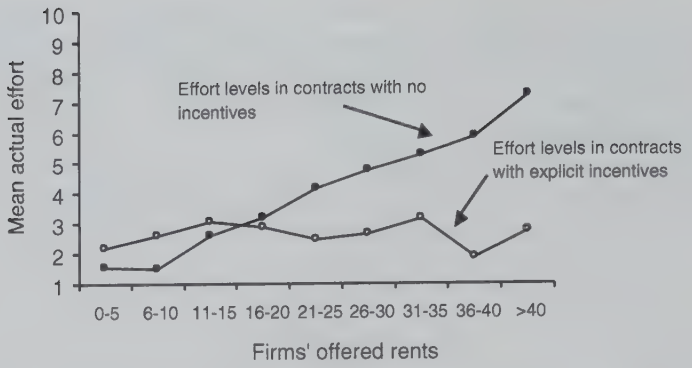



***Figure 2.6.*** *Job rents (wage-effort cost) and effort levels, with and without shirking fines. Source: Fehr and Gächter (2000b), p. 171, Figure 3; reproduced with permission of the American Economic Association.*

and firms are matched anonymously for one period at a time. If reciprocation occurs, it is purely because of a social norm or moral obligation workers feel (which firms anticipate). Furthermore, the firms' marginal profit from increases in effort $e$ is much larger than the workers' marginal cost, so a gift exchange norm creates higher wages, higher effort, and higher profit.[26]

The gift exchange account is similar to "efficiency wage theory"—wages are paid above the market-clearing level so that workers have something valuable to lose (the job rent) if they shirk and are caught by (costly) monitoring. In equilibrium, workers should therefore not shirk. The gift exchange and efficiency wages are hard to distinguish empirically with field data[27] so an experiment is ideal.

A good statistic to summarize results is the "job rent," $w - c(e)$, offered by the firm. In Fehr and Gächter (2000a) effort levels vary from 1 to 10 and firms offer a wage plus a suggested effort level $e'$. Figure 2.6 plots the job rents offered by firms ($w - c(e')$) against actual effort levels. In the treatment with no disincentive for shirking (black dots), workers respond reciprocally by choosing higher effort levels when the offered job rent is higher, just as

[26] Naturally, whether gift exchange emerges spontaneously may be sensitive to the firm's productivity gains from effort, relative to its disutility to workers. In most of the experiments by Fehr et al., the total surplus from increasing effort is large. It would not be difficult to create an experimental environment in which the joint gains from effort are lower and convergence to the low-wage/low-effort equilibrium results. Regardless, the experimental data are an existence proof that reciprocity can be very important even in competitive environments in which moral hazard is predicted.

[27] In the efficiency wage story, firms should auction off scarce jobs by charging an upfront fee, equal to the discounted value of the rent stream lucky workers will receive, but in practice this is rarely observed.

gift exchange theory predicts (even when stakes are a couple of months' wages; see Fehr and Tougareva, 1995).

Keep in mind that workers are *not* rematched with the same firms over time, so the degree of gift exchange is striking and not explained by standard concepts in game theory such as reputation building. However, the boundaries of this result remain to be thoroughly established. Hannan, Kagel, and Moser (in press) report much less gift exchange among American undergraduates, and Charness, Frechette, and Kagel (2001) find much less gift exchange when subjects are shown an explicit payoff table giving net profits for all wages and effort levels.

Another treatment in Fehr and Gächter's (2000a) study adds explicit disincentives for shirking (as in standard agency theory models). Firms can impose a preannounced fine (up to some maximum) if workers exert less effort than the required effort $e'$ and their shirking is detected (with probability 1/3). In their experiment, workers should choose $e = 4$ when faced with the maximum fine. Figure 2.6 shows that, in fact, workers exert *less* effort when threatened with fines! Their effort also does not respond reciprocally to offered job rents. Because there is sometimes shirking in the markets with fines, firms earn more in the treatment with fines, but total surplus is lower because effort levels are never very high. These classic monitor-and-fine incentive schemes reduce total output because leaving effort up to the worker risks moral hazard but actually triggers reciprocation; and, given the firm's profit function, the higher reciprocated effort creates a large surplus. Put differently, explicit incentives "crowd out" homegrown intrinsic incentives.[28]

Fehr, Gächter, and Kirchsteiger (1997) extended their design by adding a third stage in which firms can punish or reward workers after observing their efforts (in addition to probabilistic fines). In the third stage firms choose a number $p$ between 0 and 2. Their worker's job rent $(w - c(e))$ is multiplied by $p$, so that workers are rewarded with a bonus ($p > 1$) or punished by sanctions ($p < 1$). Choosing a value of $p$ further from 1 is costlier to the firm, so profit-maximizing firms will never reward or punish. In the two-stage design (with no multiplier $p$), it is impossible to incentivize workers to exert more than 10 percent of the maximal effort (since the maximum expected fine is equal to the worker disutility from exerting 10 percent

[28] Psychologists and others have argued that extrinsic incentives crowd out or extinguish intrinsic incentives in other ways. For example, children who like to color pictures do more coloring when they are suddenly paid a small piece-rate fee per picture, but they "go on strike" and reduce their coloring (to a level of output below their pre-fee rate) when the piece-rate is taken away. In my view, this is an extremely important phenomenon but has not been carefully separated from other forces. For example, in a multitasking environment, if people are suddenly rewarded for one type of output but not the other, they will naturally shift effort to what they're paid for. If this is bad for a company then it's poor incentive design.

effort). Firms ask for 30 percent effort but most workers exert less than 10 percent effort. Adding the third bonus/sanction stage improves effort and profitability dramatically. Firms ask for 90 percent effort and workers respond with about 80 percent. Firms *are* willing to punish and reward by choosing $p \neq 1$. When workers shirk ($e < e'$), they exert hardly any effort, and more than half the firms choose a low multiple (an average of 0.20). When workers work as hard as firms ask ($e = e'$), the firms pay substantial bonuses half the time (mean $p = 1.6$). As in Fehr and Gächter's (2000c) public goods games with punishment, the mere threat of punishment or bonus awards—although not credible in a world of profit-maximizing firms—is carried out often enough that it supports high effort levels and large gains from exchange.

Fehr and Falk (1999) went a step further. Their earlier experiments use a posted-offer institution reflecting how most labor markets seem to operate: Firms post wages and workers either take the jobs or don't. Fehr and Falk used a double auction institution in which both sides post offers. In a complete contract condition, the experimenter (playing the role of a regulatory institution or union contract) specified an effort level. In an incomplete contract condition, effort was unspecified. In the complete contract condition, workers ask for a wide range of wages and firms just hire those who will work for the least (since there is no need to overpay to induce worker goodwill, because workers are required to exert the demanded effort).

Figure 2.7 shows the time series of worker wage offers (dashes) and mean offers accepted by firms (connected dots) over ten periods, when offered contracts were *incomplete*. In the incomplete contract condition, workers compete fiercely for jobs by offering to work for less and less. But firms persistently hire the workers who ask for the *highest* pay. Hiring the most expensive workers actually pays because those workers usually choose higher effort levels, "repaying" the firm's "gift" of supramarginal wages. These results can be interpreted as firms knowing that it pays to share the productivity benefits of having highly motivated workers, a result often seen in field data (e.g., Krueger and Summers, 1987).

In all the experiments above, firms cannot identify a worker and recontract with the same worker repeatedly. Brown, Falk, and Fehr (2002) add the possibility of firms making private wage offers that are earmarked to specific workers. Within five periods, half the firms make private offers. If the worker they hired exerts enough effort, the firms often offer a "raise," but if effort is disappointing (relative to the firms' expectations, which are measured) the workers are "fired" (not rehired). A two-tier labor market emerges in which some workers and firms pair repeatedly, workers work hard and firms pay them well, and other firms offer a low wage to whomever is unpaired. This extension to private earmarked offers should prove a very useful paradigm

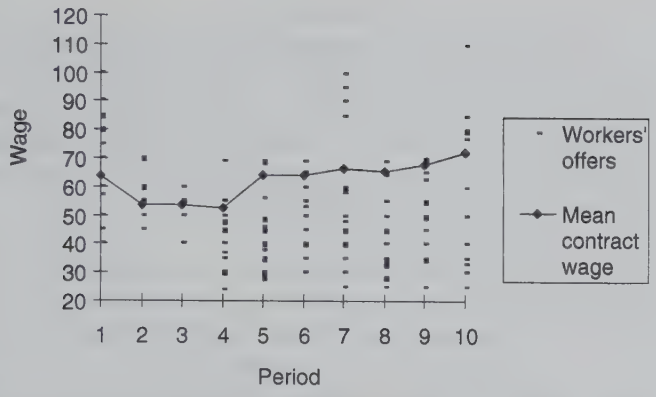



***Figure 2.7.*** *Worker wage offers (dashes) and accepted offers (connected dots) in double auction labor markets with incomplete contracts. Source: Fehr and Gächter (2000b), p. 174, Figures 4a and 4b; reproduced with permission of the American Economic Association.*

to study insider/outsider models, the stigma of unemployment, and other labor market phenomena which are often difficult to understand fully in field data.

One way to digest what gift exchange experiments tell us is to start with two caricatures of organizational design. Firms adopt rigid incentive schemes, requiring certain effort levels, punishing shirkers with fines or firing, and hiring the cheapest workers. (This scenario may describe manual labor jobs where monitoring is easy, such as service industry "McJobs.") Or, oppositely, firms eschew explicit threats and rely on worker goodwill, perhaps hiring those who ask for the *highest* pay. These scenarios loosely correspond to discussions in organizational literature contrasting the "theory X" assumption that workers are basically lazy and must be incentivized and monitored, with the "theory Y" assumption that good job design provides motivation to workers who are eager to take satisfaction from a job well done, and workers will not reliably take advantage of loopholes if they feel some moral obligation to "their" company. Economic models mostly dwell on the gloomy theory X scenario and how contracts can minimize moral hazard. The experimental results suggest that it is easy to create an experimental theory world in which moral hazard is solved by norms of reciprocation. Furthermore, in the presence of reciprocity, incentive-based solutions may do more harm than good by extinguishing the beneficial effects of reciprocation.

A careful study of successful high-tech startups suggests the "soft" gift exchange model can work well in practice. Baron, Hannan, and Burton (2001) classified 175 Silicon Valley firms into several organizational design

categories and followed their progress for several years in the 1990s. About one-fifth of the firms used a "commitment blueprint" in which the basis for recruitment and retention was a person's fit in the company's "community" and workers were controlled and coordinated by peers and company culture (rather than monitoring or formal processes). Most CEOs thought the community-type firms would be the first to go. In fact, they were the *most* successful (i.e., went public most rapidly).

**Summary:** In a pure (or blind) trust game, an Investor invests money with a Trustee she does not know, talk to, or meet again. The investment earns a return. Then a Trustee decides how much to repay. This is a classic game with moral hazard because the Trustee's repayment is not contractually enforced at all. In most experiments subjects invest about half their endowment (exceptions are the peasant–dictator games of Van Huyck et al., in which there is little trust, and in Kenya, where they invest a quarter). Repayments are generally equal to the original investment or slightly less than the original investment (trust does not quite pay). Trustees appear to repay because of a moral obligation they feel, or (put differently) a positive reciprocation of the Investor's risky action which benefits them. However, in two studies Trustees repay only slightly more than Dictators allocate (matching the amount at stake), which suggests repayments are mostly the result of altruism and are increased only a little by reciprocation.

In two-person games, trust and trustworthiness can be justified by emotional links (even if anonymous). Such links might be severely weakened when the person who repays you is not the one you trusted initially. Three studies show the opposite: There is substantial indirect reciprocity of this type (karma, in the extreme), though it is probably weaker than direct reciprocity. Any such effects suggest that "me and my group," although hardly a natural unit of analysis in game theory (contagion effects aside), could be a helpful concept for explaining evidence of substantial indirect reciprocity. An economics without concepts of group identity (cf. Akerlof and Kranton, 2000) may be missing something very important.

Gift-exchange labor markets are multiperson trust games in which firms post wage offers and workers (who outnumber firms) accept them or not. Then workers exert costly effort which firms cannot control or penalize. Self-interest predicts workers will shirk, and firms, anticipating this, offer minimum wages. In the experiments, however, firms offer large wages and workers reciprocate by working hard. Explicit incentives such as fines after probabilistic detection of shirking—the textbook solution—backfire because they undermine reciprocity. Reciprocity solves the moral hazard problem very well, and explicit incentives substitute a poor solution for a very good one.