PART III

LIMITED COGNITION: ATTENTION, PREFERENCE FORMATION, AND RISK EVALUATION

Published in *The Psychology of Economic Decisions*, 111-129. Edited by I. Brocas and J. Carillo, 2003, reproduced by permission of Oxford University Press.

7

Thinking about Attention in Games: Backward and Forward Induction

COLIN F. CAMERER AND ERIC J. JOHNSON

Behavioral economics improves economic analysis by using psychological regularity to suggest limits on rationality and self-interest (e.g. Camerer and Loewenstein 2003). Expressing these regularities in formal terms permits productive theorizing, suggests new experiments, can contribute to psychology, and can be used to shape economic policies which make normal people better off.

The crucial leap in behavioral economics is taking the details of thinking seriously, rather than settling for the conventional apology that models are only "as if" representations of the output of a neural "black box." Thinking seriously about cognitive detail, and formalizing those ideas, has become more fashionable in economics recently. Models of intertemporal choice seek to understand precisely how people think about the future (see Frederick, Loewenstein, and O'Donoghue 2002). Theories of social preference formalize how people feel about allocations of payoffs to themselves and others (Fehr and Gachter 2000; Camerer 2003: ch. 2). Another major innovation in economics is the observation that people think differently about gains and losses, which implies that the way outcomes are described or "framed" can affect which choices are made (Tversky and Kahneman 1991; Rabin 2000).

Strategic interactions among many players, or "games," are of special interest for behavioral economics because even games which are very simple representations of naturally occurring situations can give rise to equilibria that strain computational ability. As a result, games are a domain in which models of limited thinking are quite likely to provide better predictions than equilibrium theories. Indeed, behavioral game theory researchers have developed precise models of how deeply people are thinking about games which are as general as equilibrium theories, typically more accurate (and often more precise as well; Nagel 1995; Stahl and Wilson 1995; Costa-Gomes, Crawford, and Broseta 2001; Camerer 2003: ch. 5; Camerer, Ho, and Chong 2003*a,b*).

AQ: Please check if this should be Nagel 1999 as in reference list.

AQ: Please note that Camerer et al. 2003b is not listed.

The forward induction data were collected in collaboration with Gerard Cachon.

This chapter discusses how recording the information people seek in games can permit sharp inference about what goes on inside the black box (the brain), and can therefore inform theories about why people behave the way they do. Until recently, experimental economics was strongly influenced by the tradition of identifying unobserved (latent) utilities with choices, or revealed preferences. This tradition focuses attention on only one type of data from a decision—the observed choice. Looking only at choices is scientifically efficient because many other aspects of cognition *can* be measured (as we show below); and these other measures can be used to distinguish between alternative theories which can account for observed choices equally well. Indeed, as economists have become more interested in evolutionary and behavioral approaches, the number of theoretical alternative accounts has increased. The "process data" described in this chapter help by providing a richer set of data which allows us to sort through alternative explanations more quickly.

AQ: Please check if these asterisks should be Quotes

Measuring information search improves scientific efficiency in two ways:

Process data facilitates the evaluation of multiple theories. In early experiments in game theory, the fundamental question was whether players chose equilibrium strategies (and when alternative equilibrium concepts made different predictions, which were most accurate). From this relatively well proscribed set of alternatives, a game theorist must now consider many more possibilities, including concerns about fairness and limited cognition. Studying thinking allows us to explore this space of alternative explanations more efficiently by providing additional measures and constraints within the same experimental session. For example, in bargaining games fairness theories require that players look at the payoffs of others, even if those payoffs are strategically irrelevant. Theories of limited cognition can therefore be tested by looking at what information players *do not* examine.

Process data can help us understand heterogeneity among individuals. While a single experiment might provide a dozen observed choices for an individual, augmenting that study with process data can greatly increase our ability to understand how individuals differ, when quite different processes lead to the same observed choices. As a result, we can identify which "type" an individual is more quickly if we can watch the inputs to thinking. Understanding heterogeneity is important because it is well-established that how different types interact is important in determining system-wide behavior.[1]

This chapter gives two concrete examples of the advantages of using process data to study cognition in games, and also mention the caveats of using process data. We start by offering a quick overview of process measures, then provide a quick review of several studies that have applied process data to games. Finally, we close by discussing the future of process data in economics.

[1] The most familiar examples is models of repeated games, in which a small percentage of types with "unusual" behavior can profoundly affect what other players do.

1. VISITING THE THINKING FACTORY

There are many ways to study the psychological processes underlying cognition. We lump the results of these different methods together under the term "process data," meaning the observe the flow of information used in service to cognition. All the methods share one characteristic: They observe how people acquire information. By observing the arrival and use of information, we can infer something about the nature of underlying cognition. We are *not* talking about peoples' self-reports about their own thinking. While self-reports are sometimes insightful, there are well-known problems with introspective access to cognition which limit what can be learned from just asking people what they were thinking (e.g. Nisbett and Wilson 1977).

Johnson et al. (2002) introduced the analogy of the human brain as a "thinking factory." One can make educated guesses about a factory's production process by observing the flow of inputs into the factory, the length of processing time before new inputs enter the factory, and the factory's final output without ever looking inside the factory. Lawyers studying jury decisions (whose deliberations cannot be directly observed) and intelligence agencies make inferences like this routinely. Similarly, the order in which subjects gather information, and how long they use information before getting more information, can be used to make educated guesses about how subjects are thinking (i.e. the production process in the neural "thinking factory"). If information is not acquired, it cannot be used, just as a car factory which never receives shipments of rust proofing chemicals cannot possibly be rust proofing its cars.

An analogy can be made to revealed preference. Choices between objects (such as commodity bundles or gambles) reveal unobservable preferences. Similarly, studies of industrial output reveal unobserved production functions. Asking subjects to "choose" information is asking them to reveal their preference for information, and this can provide clues to unobservable thinking patterns.

Psychologists and consumer researchers have employed many process tracing techniques, including recording people "talking aloud" concurrently while thinking (Ericsson and Simon 1980), watching the physical retrieval of information required to solve a problem (Payne 1976; Jacoby et al. 1985), or asking people what information they would use.

The basic idea behind all process tracing techniques is illustrated by recording eye fixations, because it underlies most visual information acquisition. It surprises most people to learn that our perception of the visual world is, in part, illusory. While we believe that we see a continuous stream of visual input, the eye is more like a still camera taking snapshots than a movie camera filming continuous scenes. For almost all vision that requires significant acuity, the eye makes a series of jumps or *saccades* between pauses, called *fixations*. During the 100 ms or so required for a saccadic movement, vision is actually suppressed. Brief lights flashed during this eye movement are not seen. The brain takes the sensory output of these discrete snapshots each lasting around 0.2 to 5 s and creates the perception that we see a continuous, uninterrupted view. The beauty of eye fixation recording is that it is exactly what goes on in the real world: All high level cognition, such as reading, uses eye fixations, and eye fixations have been quite useful in studying reading and similar tasks (Just and Carpenter 1976). Eye fixations have also been used to study choice (Russo and Rosen 1974). But recording eye fixations (typically by recording the direction of eye gaze relative to stimuli spaced sufficiently apart on a screen) is fairly expensive and somewhat uncomfortable for participants. It is also not easily adapted to group settings.

Many researchers now employ computer-based information acquisition as an analog for recording eye fixations. These methods display information hidden in boxes on a computer screen and observe a person as they "acquire" or "look up" information using some kind of pointing device like a mouse. People navigate through a task by moving the mouse and clicking on boxes which reveal information they want to know.

Tracking "clicks" is now a major activity for Internet web sites, but computer-based information acquisition has a long history that predates the web. It has been used extensively to study individual decision-making and consumer choice. Studies of decision-making include the choice of gambles (Payne and Braunstein 1978), the setting of reservation prices for gambles using different response modes (Johnson and Schkade 1989), the effects of time pressure on choice (Payne, Bettman, and Johnson 1993; Reiskamp and Hoffrage 2000) and the study of preference reversals (Schkade and Johnson 1989). Studies of consumer choice include Jacoby et al. (1985).

Au: Please check if insertion of year is OK here?

Computerized methods offer several advantages over verbal self-reports, eye movements, and older methods such as having subjects request information displayed on file cards. Using a mouse reduces the impact of the measurement procedure upon the underlying process. Computer-based methods also easily adapted to group situations, unlike verbal protocols and eye movements. The recording of information acquisition is automatic and unobtrusive, so experiments can be run on networked computers without participants' awareness. An important question is whether recording process data in these ways changes the process (a la the Heisenberg uncertainty principle in physics—does the act of experimental observation change the process being observed?). The answer is generally "No" but we defer further discussion to the end of this chapter.

2. THINKING BACKWARDS

Consider an alternating offer sequential bargaining game (Ståhl 1972; Rubinstein 1982). This study (Camerer et al. 1993; Johnson et al. 2002) illustrates the use of process data to (*a*) distinguish among alternative theories and (*b*) to examine individual differences.

Au: Please check if the year for Camerer et al. should be 1994

Au: Please check accent å in Ståhl 1972 has been changed to match the reference list.

In this game, two players, 1 and 2, bargain over a pie that shrinks in value in each of three periods (reflecting discounting due to impatience). The pie is worth about \$5 in the first period, \$2.50 in the second period, \$1.25 in the third period, and nothing after that. Starting with player 1, the players alternate making offers, which the responding player can accept or reject. If an offer is rejected, the pie shrinks, and the player who rejected it then makes a counteroffer.

Play is not well predicted by equilibrium: If players are purely self-interested (and believe others are too)[2], perfect equilibrium divisions can be derived by backward induction, starting with the third-period division and working backward. The perfect equilibrium is (approximately)[3] for player 1 to offer \$1.25 and keep the rest for himself. In experiments, subjects typically offered something between the \$1.25 equilibrium and \$2.50 (which is an equal split of the first-round pie). Offers average around \$2.00. Lower offers, including offers near the equilibrium prediction of \$1.25, are frequently rejected.

At least two theories offer possible explanations. The first, limited computation (LC), explain observed departures from perfect equilibrium by suggesting that players do not reason game-theoretically, and hence do not initially understand how the structure of the game conveys bargaining power. The second class of theories uses equilibrium social preferences to account for departures from perfect equilibrium. They posit a "social utility" for others' payoffs or differences in payoffs (Loewenstein, Thompson, and Bazerman 1989; Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Charness and Rabin 2002; Camerer 2003: ch. 2), a preference for reciprocating fairness and unfairness (Rabin 1993), or an unobserved component of payoffs which appears like noise or a mistake to an outside observer (McKelvey and Palfrey 1995, 1998; Goeree and Holt 2001; Weizsacker, in press).

Au: Please update the reference of Weizsacker.

Of course, these two theories are not mutually exclusive, and people may well differ in the degree they believe in fairness or have limited computational abilities. Clean predictions based on choices alone are ambiguous. Table 7.1 shows four theories which combine elements of fairness and limited computation. We describe each combination briefly.

Au: Please check if the expansion introduced for GT is correct.

If players care only about their own payoffs, and can compute perfect equilibrium, then game theory (GT) predicts that players will offer \$1.25.

[2] Of course, the perfect equilibrium prediction in an n-stage game requires n levels of iterated belief in self-interest. For example, in our game player 1 offers \$1.25 only if she is self-interested, believes 2 is (which guarantees a minimal ultimatum offer in the third round), and believes 2 believes she (1) is (which guarantees a \$1.25 offer to 1 in round 2, leading player 1 to offer \$1.26 in round 1).

[3] In the third round, player 1 offers \$0.01 or nothing (depending on whether player 2 will accept nothing) and earns \$1.24–1.25. In the second round, player 2 must offer \$1.24–1.26 to player 1, leaving \$1.24–1.26 to herself. In the first round, player 1 must therefore offer \$1.24–1.27 to player 2. For simplicity, we refer to this range of perfect equilibria as "the" equilibrium at \$1.25, recognizing that the equilibrium is not unique, and is subgame perfect rather than only Nash (*every* offer is consistent with Nash). Note that a short cut to calculating equilibria in games of this form is that the first player earns the sum of the pie sizes in odd-numbered rounds (\$5.00 + 1.25), minus the pies values in even-numbered rounds (−\$2.50). Alternatively, the first player should offer the third pie minus the second pie to the second player, which does not even require her to know the first-period pie size.

Table 7.1. *Model predictions*

Theory	Treatments: Bargaining with		
	Other subjects	Robots	Robots and instruction
GT	\$1.25	\$1.25	\$1.25
GT-fairness	[1.25, 2.50]	1.25	1.25
LC	p in [0, 2.50]	p	1.25
LC-fairness	$p_1 \geq p$	$p_2 < p_1$	1.25
Mean offer	\$2.11	\$1.84	\$1.22

Now suppose players can compute perfect equilibrium, but they either care about fairness or believe others do. This combination of "game theoretic with fairness" (GT-F) suggests that offers will lie in an interval between the GT prediction and the equal-split point, [\$1.25, \$2.50]. If players do not reason game-theoretically and not initially understand how the structure of the game creates bargaining power, they might be labeled as possessing Limited Computation. Such a "limited computation" (LC) theory, predicted offers could lie anywhere; if we assume an upper bound at the equal-split point \$2.50, we predict an offer p in the range [0, 2.50]. If players have limited computation *and* are concerned with fairness, we predict offers will be more generous than in the LC case, $p_1 \geq p$. Note that under these conditions, only the GT prediction seems inconsistent with the existing data, and our observed offers and rejections.

Clearly, additional help might be useful in understanding the relative role of these two accounts. In Johnson et al. (2002) we gathered process tracing data using the MouseLab computer-based information monitoring system (Johnson et al. 1991). This system displayed the parameters of the problem behind the labeled boxes. Players acquired the information by moving the cursor into the box. When the cursor left the box, it reverted to a blank square (Figure 7.1).

Au: Please check citation of figure 7.1 inserted here is OK.

Software records the entrance and exit time for each box, typically to the 1/60th of a second. Typically these data are characterized by a dense series of acquisitions with durations much like those generated by eye fixation recording, ranging in length from 200 ms to a few seconds in length. Consistent with the notion that this is a relatively effortless and costless response, there are many occasions when subjects open the same box more than once, showing that it is easier to move the mouse than memorize something.[4] Figure 7.2 shows the data, both in terms of number of acquisitions and in terms of the distribution of total time. Figure 7.2 also demonstrates that using process tracing data represents a challenge in data presentation: How can we show mean looking times and

[4] There are exceptions. We had the opportunity to demonstrate this design to a well-known game theorist, who used the mouse (with two hands) and wrote the numbers down on a piece of paper. Fortunately for us, such behavior is exceedingly rare: To date it has been confined to Nobel Prize winners.

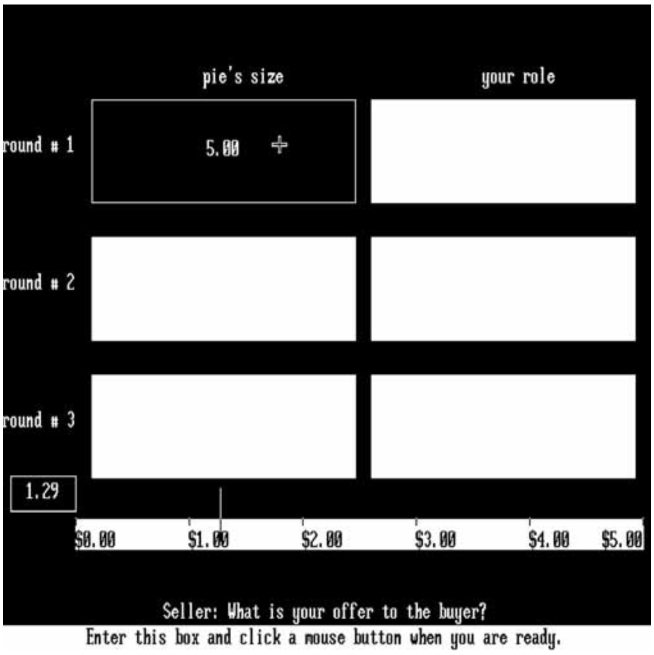

Figure 7.1. *Computer display, alternating offer game*

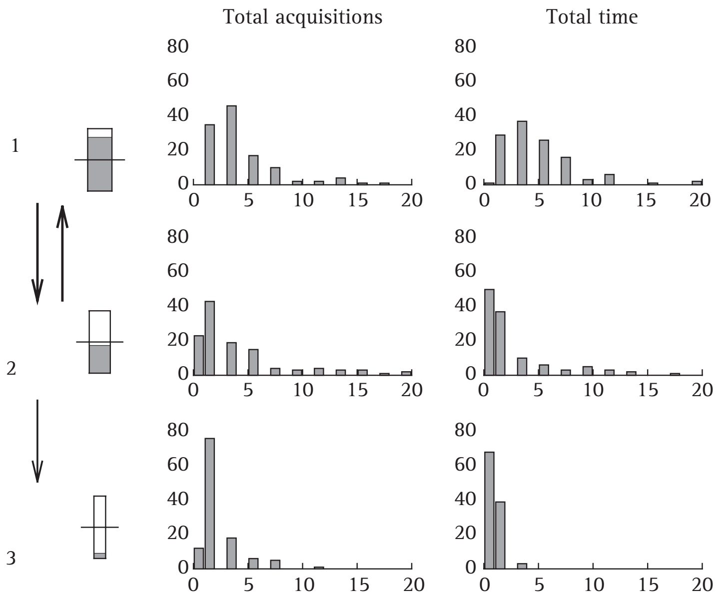

Figure 7.2. *Distribution of acquisitions, and transitions by round*

numbers of look-ups, as well as transitions from cell to cell? The "icon graph" at the left of Figure 7.2 is one method to show all these data: For each box (i.e. each amount to be divided in each round), the width of the bars is proportional to the number of acquisitions, the height is proportional to the total looking time, and the black arrows to the left of the figure show the frequency of transitions. The thickness of the arrows is proportional to the frequency, with fewer than one transition on average omitted.[5]

Like most studies, we found a fairly broad range of offers in the first round from around the equilibrium of \$1.25 to some just above the midpoint of \$2.50. Based upon choice data alone, we were unable to distinguish between theories. The process tracing data, shown in Figure 7.2 seems more compatible with a limited cognition view: Many subjects do not look at the subsequent rounds: 19 percent do not open the second box, and 10 percent do not open the third. Both pieces of information would be necessary to calculate the equilibrium.

Yet, there seems to be great heterogeneity in offers and search. Without process data, the range of offers might be duly noted and remain unexplained. To help us understand this heterogeneity, however we posited three types of information acquisition, roughly corresponding to levels of look ahead: Level 0 considered only the current payoff in Round 1, Level 1 looked at this round payoff and that in the next round, while Equilibrium players must consider all three rounds.

Types are classified by identifying which box was looked at longest in the first bargaining round. This is simple, and allows us to test if the offers are indeed related to search. Figure 7.3 displays the results for the process tracing data and offers, aggregating across our first two studies. The top of the figure displays an icon graph that is very much as predicted: Those trials which are characterized as exhibiting "Level 0 look ahead" spend most of their time with the first round payoff, spend very little time looking at other payoffs, and make very few comparisons between subsequent payoffs. Their mean offers are \$2.07. Level 1 cognition is characterized by increased time spent on the second round payoff, and produces a mean offer of \$1.71. Finally, those trials which are characterized by the most acquisitions of the last round payoff look much like we expect from backward induction: Minimal examination of the first round payoff, concentration on the second and third round payoff with many comparisons, and a mean offer of \$1.44. While there is substantial dispersion in the offers produced within each information-lookup type category, the differences in the offer distributions across the three columns show that it is possible to statistically predict what a player will offer from what information they look at most often. The big loser among the alternative theories is the idea that players make nonequilibrium

[5] The skewness of these data is typical. For analysis, it is often useful to either use nonparametric methods, which often lack statistical power, or to analyze the data with parametric techniques after a log transform $x' - log(x + c)$ where c is a small constant, typically 0.5, used to prevent taking the log of 0. Means are reported after an inverse transformation.

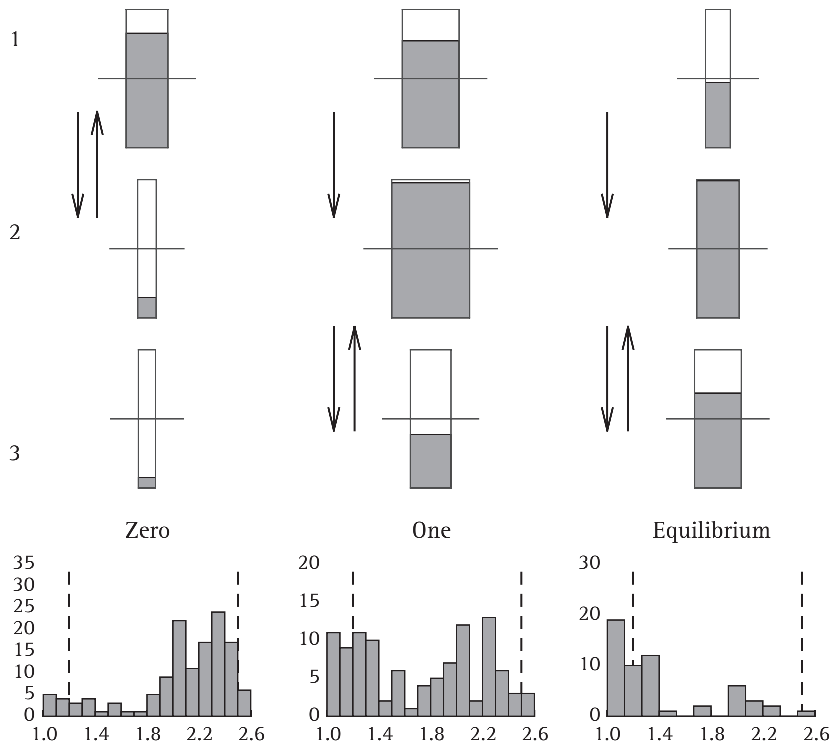

Figure 7.3. *Search and offers by type*

offers because they do equilibrium computations but adjust for fairness concerns of others ("GT-fairness"), since the most fair offers come from subjects who simply do not look ahead. Furthermore, in one experimental condition subjects bargain against a computerized opponent who is programmed to maximize its own payoff (and to expect that other subjects will too). In that condition, most subjects do not offer \$1.25, which they would if they were computing money-maximizing offers optimally; although they learn to do so quite rapidly with a little instruction in backward induction.

This study shows that process data allows us to both (*a*) increase the set of strategies considered (to include steps of thinking) and (*b*) provide a better account of heterogeneity in these data.

3. THINKING FORWARDS

Figure 7.4 shows a game we used to study forward induction. Player 1 can choose the outside option and guarantee a payoff of 11, or play a simultaneous-move "battle of the sexes" subgame, choosing up or down. What would you do?

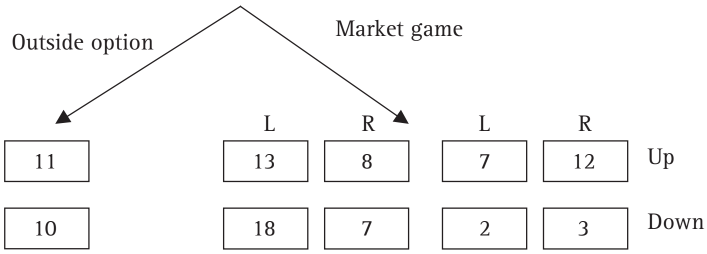

Figure 7.4. *Forward induction game*

A rational player A plays the market subgame and then plays Down,[6] expecting player 2 to reason as follows:

> Player 1 could have had a payoff of 11. Instead, she entered the game, signaling that she intends to get more than 11. The only equilibrium which yields more is (Down, Left). Since I have deduced that she will move Down I should move Left.

This clever argument is called "forward induction" (Kohlberg and Mertens 1986; van Damme 1989): Choices at current nodes have implications for play in future nodes. To do forward induction, player 2 must look player 1's choices, ones which will not be played and which are not choices for player 2. Player 2 must reason that those payoffs may offer clues about what player 1 intends to do.

We ran three sessions of games similar to that in Figure 7.4, with 14 subjects each. (The games differed in player A's option payoff, either 14 or 11, and player B's option payoff, either 6, 10, or 16.) Contrary to forward induction, player 1 subjects chose the outside option more than half (55 percent) of the time. (In a structurally similar game, Schotter, Weigelt, and Wilson 1994, observed 70 percent choice of the outside option; Camerer (2003), chapters 5 and 7, summarizes other studies showing weak support for forward induction). When player A's did choose to play the matrix game, they chose the bottom row 95 percent of the time; so the player A's who anticipated that player B's would do forward induction were betting heavily by choosing the bottom row almost all the time. In addition, B's played left about 75 percent of the time. Furthermore, there was little change across twelve repetitions, indicating no detectable learning.

AQ: Please check if box S11 should be box $11 as in fig 7.5

Process data provide a partial answer to why the forward induction was rare. Figure 7.5 presents an icon graph of player B's looking time. Note that there are very few lookups of acquisition of either outside option payoff—the box S11 icon is skinny and the associated looking time is low (little shaded area)—although player B has to look at player A's option payoff (box S11) to make a forward induction inference. But note that player B's do not think that

[6] More formally, the subgame has two pure-strategy Nash equilibria, (T, r) and (B, l), and a mixed-strategy equilibrium $(0.5T + 0.5B, 0.5r + 0.5l)$ (with expected utilities $(10, 7.5)$). The entire game has Nash equilibria in which player 1 moves L because she thinks subgame play will give the player less than 14. Of those, the equilibria (LT, r) and $(L(0.5T + 0.5b), 0.5r + 0.5l)$ are subgame perfect too. The perfect, Nash equilibrium (RB, l) is the unique "stable" equilibrium which satisfies forward induction.

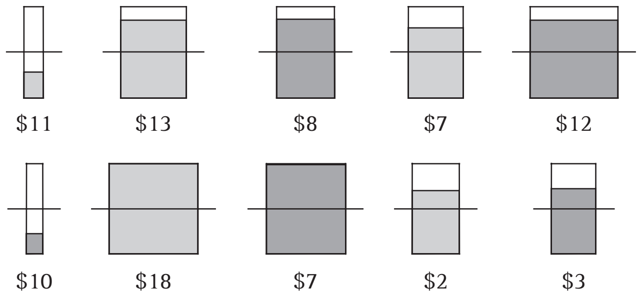

Figure 7.5. *Icon graph, player B*

player A's payoff are irrelevant; in fact, they spend considerable time looking at player A's payoffs in the matrix subgame. It is just the outside option payoffs of both players that are neglected by player B's. This is particularly interesting because both players alternated roles. It is also interesting because player A's *are* affected by the outside option.

Comparing behavior when payoffs change also helps clarify what players are doing. Recall that player A's option payoff is either 11 or 14, in different treatments. In both cases the option payoff is above the "bad" pure equilibrium payoff for player A (which is 7) and also above the expected mixed-strategy equilibrium payoff (which is 10). Forward induction points to the "market, (bottom, left)" equilibrium in both cases. But when the option payoff is 14, the "market, (bottom, left)" equilibrium can also be reached by three steps of deletion of dominated strategies: If B's think that A's will never violate dominance, they think the A's will never play "market, top"; eliminating that dominated strategy, B's will always choose "left"; if A's believe B's will play that strategy, A's will play "market, bottom." Thus, if A's think that B's think that A's will obey dominance, A's should play "market, bottom." If B's think A's will obey dominance the B's will play "left."

Thus, when A's option payoff is 14, applying dominance iteratively leads to the "market, (bottom, left)" outcome even without forward induction reasoning. One might, therefore, expect that the forward induction outcome is more frequent when A's outside option payoff is higher, 14 rather than 11 (since *both* forward induction and iterated reasoning predict that outcome when A's option payoff is 14). Intuitively, in rational analyses the higher option payoff for A gives the B players more reason to believe that A will play the market game, and more reason—if they are iteratedly rational—to infer that they should play "left." This, in turn, draws strategically-minded A's into the market game. In fact, the opposite is true: When the option payoff was 14, players A entered the market game only 38 percent of the time; when the option payoff was 11 they entered the market game 57 percent of the time. So changing A's option payoff had the *opposite* effect of what a rational analysis would predict.

The fact that more player A's chose the option when its payoff was higher (14 rather than 11) is consistent with a model in which players are limited in their strategic thinking, but are not sure other are rational (e.g. Camerer, Ho, and Chong 2003*b*). Crawford (2003) put a similar analysis to work in a careful study of deception.

About 10 years ago, we created some behavioral game theory around this concept—especially the idea that players treated choices of other players as gambles, rather than as the result of a delicate (iterated) reasoning process. At that time, we did not know precisely how to model this idea, and rationality-based game theory was so roaringly productive that buyers for a cognitively plausible alternative were scarce. It is heartening that theorists eventually accepted the mathematical challenge of modeling the limits on rationality that we envisioned, and are now relatively hungry for data to guide their theorizing.

4. THINKING NORMALLY

These studies explore whether players look backward and forward in "extensive-form" games that take place in several stages. Inspired by our earlier studies, Costa-Gomes, Crawford, and Broseta (2001) and Costa-Gomes and Crawford (2002) used MouseLab to answer a more basic question in game theory: How strategically do players think in "normal form" (matrix) games?

Au: Please check if change is O.K.

Their ambitious study used eighteen different games, with 2–4 strategies for each player. The games were carefully designed so that players using different decision rules would pick different strategies, and different rules imply different patterns of information acquisition. Assuming each player has a single strategy or "type," they then use a sophisticated Bayesian procedure (cf. Harless and Camerer 1994; El-Gamal and Grether 1995) to infer which type a player is most likely to be, given the player's choices and information acquisition across the eighteen games.

Some of the decision rules they consider do not use sophisticated strategic thinking about what others will do. These rules are "naïve" or level 1 (L1) (choose strategies with the highest average payoff, averaging all payoffs equally; "optimistic" or maximax (choose the strategy with the highest possible payoff); "pessimistic" or maximin (choose the strategy with the highest possible minimal payoff); and "altruistic" (maximize the sum of the two players' payoffs).

There are also five strategic types: "L2" (best-respond to L1); "D1" which does one round of deleting dominated decisions, then best-responds to a uniform prior on the remaining decisions; "D2," which does two rounds of deletion then best-responds to remaining strategies; "sophisticated" who guess accurately the proportion of the population that chooses each nonequilibrium strategy;[7] and "Nash" (equilibrium).

[7] Importantly, in the first version of their paper the L2, D1, and D2 types were not included in the analysis, and their procedure classified nearly half of the subjects as "sophisticated." When these types are included the fraction of estimated "sophisticates" falls almost to zero, which shows that specifying a rich space of decision rules is important to draw the right conclusions.

As in our work, subjects participated in baseline and "open boxes" conditions, to see whether hiding payoffs behind boxes limited strategic thinking (it did, but only a little). They also trained subjects to use the various decision rules (and rewarded them for picking the strategies those decision rules should pick) to calibrate what information acquisition looks like for various decision rules.

The percentages of untrained subjects choosing equilibria were 90, 65, and 15 percent when equilibria require 1, 2, and 3 levels of iterated dominance, respectively, which is consistent with much other evidence that people do only one or two steps of strategic reasoning (e.g. Nagel 1999; Camerer 2003: ch. 5; Au: Please indicate if it is 2003 'a' or 'b' Camerer, Ho, and Chong 2003). However, trained subjects find equilibria almost perfectly (90–100 percent of the time). Like our finding that human subjects exposed to backward induction quickly make equilibrium offers to robot subjects, the powerful effects of training show that strategic thinking is not "computationally difficult" *per se*; it just requires decision rules which are not unintuitive, but easily taught. Strategic thinking is not like weight lifting or dunking a basketball, where performance is constrained by physical limits. Instead, strategic thinking is more like learning to windsurf or ski, which require people to learn skills which are unnatural and overcome natural instincts (overcoming the reluctance to lean forward going downhill and backward toward the water). Learning backward induction requires overcoming the intuition that future subgames which are unlikely to be played can be ignored; and learning iterated reasoning requires looking as much at another player's payoffs as one's own.

Since different decision rules point to different strategies in each game, looking at a player's strategy choices can classify the player according to which decision rule they use most often (cf. Harless and Camerer 1994, 1995; El-Gamal and Grether 1995).

Table 7.2 shows the fraction of subjects estimated to be using each of the decision rules, when the estimation is doing using either the decisions alone,

Table 7.2. *Estimated frequencies of various decision rules and their expected payoffs*

Decision rule	Information used to estimate frequencies		Expected payoff
	Decisions	Decisions and search	
Altruistic	0.089	0.022	17.11
Pessimistic	0.000	0.045	20.93
Naïve (L1)	0.227	0.448	21.38
Optimistic	0.000	0.022	21.38
L2	0.442	0.441	24.87
D1	0.195	0.000	24.13
D2	0.000	0.000	23.95
Equilibrium	0.052	0.000	24.19
Sophisticated	0.000	0.022	24.93

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

or the decisions plus information search data. The table makes three important points:

First, about 90 percent of the subjects appear to use simple rules (L1, L2) which are only minimally strategic (i.e. a player A does not assume that B is reasoning about A). Hardly any consistently choose equilibrium.

Second, if CGCB had done their study using only decision data, without process data, they would have drawn the wrong conclusion. Here is why: When only decisions are used to classify players, about 20 percent of players are assumed to use each of the decision rules L1 and D1. But when information search is used as well, the 19 percent originally classified as D1 shrinks to zero, and the percentage classified as L1 roughly doubles. The reason for this is that the D1 and L1 types often make the same decisions, but the D1 types must look at other players' payoffs (to see whether other players have dominated strategies that can logically be eliminated), which L1 types do not. Most of these players look at others' payoffs rarely, so the process data show it is statistically much more likely that they are L1's than D1's (despite the fact that the predicted choices of those types often coincide). The process data, therefore, overturns an incorrect conclusion drawn from the decision-only data.

Au: Please check not listed in reference.

Third, there appears to be a sensible cost-benefit tradeoff between complexity of rules and their payoffs which explains why most subjects use simple rules (cf. Johnson and Payne 1985). The second column of Table 7.2 shows the expected dollar payoff to subjects using each type of rule (if they were randomly matched with every other subject, given how subjects actually behaved). The L1 and optimistic rules make a little more than \$21 and do not require looking at the other player's payoffs at all. The L2 and D* rules earn about 10 percent more, about \$24, and L2 actually earns *more* than equilibrium! This might seem anomalous since it is easy to confuse "equilibrium" and "optimal." But equilibrium strategies are only universally optimal—that is, better than any other strategy—if they are dominant strategies. But generally, an equilibrium strategy can be a bad response to disequilibrium moves by others, and simpler rules like L2 appear to be more robust. Note also that since the sophisticated types best respond to the actual mixture of decision rules being used and by definition they have the highest expected earnings. However, they earn only *\$0.06* less than the L2's, which means that being fully sophisticated is only slightly more profitable than doing two steps of strategic thinking.

5. CONCLUSIONS

In this chapter we presented three examples of how process data can be used to understand how people play sequential and matrix games. In all three studies, motivated, intelligent subjects behave sensibly, but do not exhibit the extent of strategic reasoning which is commonly assumed when game theory is applied to understand auctions, industrial pricing, political maneuvering, incentive design, and so forth.

In the sequential bargaining studies, subjects rarely look ahead to future nodes which they believe (correctly) are unlikely to be reached. In the forward induction game, most second-mover players do not look back at the first-mover's possible payoffs (even though she might use the foregone payoff to infer something about what the first-mover will hope to get in the future, and hence what she might do). In the matrix games, most players appear to do one or two steps of strategic thinking.

6. CAVEATS

It is sensible to wonder whether the experimental procedure measuring process, by requiring a player to move a mouse to open boxes on a screen display, biases the results in some way. These concerns have been raised many times and addressed by further experiments:

Does process tracing change the underlying process, is it reactive? Within decision research, there is a significant literature replicating many standard phenomena established with pencil-and-paper methods, such as preference reversals, using computer based techniques (Payne, Bettman, and Johnson 1993), so the method of measurement does not change the observed result in these cases. Researchers have also directly compared behavior in games when players *do not* have to open boxes, because they appear on the screen uncovered, with MouseLab trials where a mouse is used to open boxes. Costa-Gomes, Crawford, and Broseta (2001) presented one session of their normal form games without covering up the payoffs and noted no difference in observed choices. In our work on sequential bargaining, we also ran a control session with open boxes, and the offers were within $0.01 of offers when boxes were closed and a mouse was used. Lohse and Johnson (1996) have compared eye fixation recording with MouseLab and found only small differences (typically when the display was large, probably due to limited peripheral vision).

Can process tracing data really tell us about underlying decision processes? Will it know a game theorist when it meets one? Costa-Gomes and Crawford (2002) and Johnson et al. (2002) ran controls where subjects were trained to follow game theoretic principles, to see what information acquisition by "game theorists" would look like. In our work, we ran several groups who had been trained to calculate the backward induction equilibrium. Their process was quite different from untrained subjects bargaining with each other (or with computerized opponents)—namely, the trained subjects opened the third and second round payoff boxes much more often, and made many more backward transitions third round payoff box to earlier-round boxes. A similar control has been run for normal-form games by Costa-Gomes et al. with similar results. In the decision literature, Bettman, Johnson, and Payne (1990) reports success training subjects to follow various choice strategies.

How important is the position of payoffs in the display? In most western languages, people read from left to right and top to bottom. One might well be concerned that this left–right, top–down bias is reflected in which boxes are opened first. To measure this bias, we ran a group of subjects where the order of the payoffs were reversed from top to bottom (i.e. the third round payoff was at the top and the first round at the bottom). The results looked like an inverted version of Figure 7.2, and replicated the results we discuss below.

Together, these results indicate that forcing players to open boxes to gather information does *not* produce important changes in cognition or eventual choices.

7. AVOIDING PITFALLS

There are two potential pitfalls in collecting and analyzing process data. One is that it is crucial to minimize memorization so that we can rely on players' constant reacquisition of information as they need it (rather than remembering). This design desideratum favors complex displays and subjects who are facile in using mice or other technologies. The fact that subjects typically open the same box more than once in a trial—often many times—suggests it just as easy for them to move a mouse as it is to remember.

The second pitfall is that observations of decisions trickle out, like a dripping faucet, while process data are like a fire hose of information. Since there are *so many* data (and multiple measures), it is difficult to see what is going on without having a strong theory that tells you exactly what patterns to look for. A strong theory also tells you how to design the experiment. For example, Costa-Gomes et al. carefully chose games in which different decision rules led to different choices and to different patterns of information acquisition.

8. THE FUTURE

Studying thinking is hard work. By definition, thinking is an internal event and is unobservable. There may be great promise in the new technologies in neuroscience such as functional magnetic resonance imaging (fMRI), but currently there are significant challenges in these endeavors: Gathering data is costly, the interpretation requires significant knowledge of the brain and its structures, and the data has mediocre limited time resolution. In fact, progress in neuroscience has been rapid largely because results are only accepted when different methods (animal studies, lesion patients, physiological measures, fMRI) corroborate one another (see Camerer, Loewenstein, and Prelec 2003).

To economic theorists, optimization is a familiar hammer which makes everything look like a nail. As a result, economists will find it irresistible to build models in which the choice of decision rules and information can be "reverse engineered" as optimal solutions to decision making under constraint (e.g. Gabaix and Laibson 2000). Optimizing theories have proved very useful, and "optimally suboptimal" theories of bounded rationality can be useful too, provided they stick close to the details of what is known about cognitive mechanisms, or make sharp empirical implications. It is easy for such exercises to degenerate into "just so" stories which posit a very special set of premises under which a behavior is proved to be optimal. However, it is usually difficult to prove that the posited set of premises are the *only* ones under which the behavior is optimal (i.e. it is easier to prove that the premises are sufficient than that they are necessary). As a result, it is legitimate to judge whether the premises are reasonable using other sources of data. To be taken seriously, such theories should therefore take empirical constraint seriously on either the front end of the modeling exercise (grounding the basic features of the model firmly in empirical regularity) or on the back end (requiring a sharp implication which would falsify the model and is worth testing), or both.

REFERENCES

Bettman, J. R., Johnson, E. J., and Payne, J. W. (1990), "A componential analysis of cognitive effort in choice." *Organizational Behavior & Human Decision Processes*, 45(1): 111–39.

Bolton, Gary E. and Ockenfels, Axel (2000), "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, 90: 166–93.

Camerer, Colin F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton: Princeton University Press).

Camerer, C., Ho, T. H., and Chong, K. (2003*a*), "Models of *thinking, learning and teaching* in games." *American Economic Review Papers and Proceedings*.

—— Ho, T. H., and Chong, K. (2003*b*), "A cognitive hierarchy analysis of thinking in one-shot games." www.hss.caltech.edu/∼camerer/thinking2002.pdf.

Au: Please provide complete publisher details for reference Camerer et al. 2003a.

—— Johnson, Eric, Rymon, Talia, and Sen, Sankar (1994), "Cognition and framing in sequential bargaining for gains and losses." In K. Binmore, A. Kirman, and P. Tani (eds.), *Frontiers of Game Theory*, pp. 27–47 (Cambridge, MA: MIT Press).

—— and Loewenstein, George (2003), "Behavioral economics: Past, present, future." In C. F. Camerer, G. Loewenstein, and M. Rabin (eds.), *Advances in Behavioral Economics* (Russell Sage Foundation/Princeton University Press.)

Au: Please provide place of publication.

—— Loewenstein, George, and Prelec, Drazen. "Neural economics: How neuroscience can inform economics." Unpublished, http://hss.caltech.edu/∼camerer/camerer.htm.

Charness, Gary and Rabin, Mathew (2002), "Understanding social preferences with simple tests." *Quarterly Journal of Economics*, 117: 817–69.

Costa-Gomes, M., Crawford, V., and Broseta, B. (2001), "Cognition and behavior in normal-form games: An experimental study." *Econometrica*, 69: 1193–235.

Costa-Gomes, Miguel A. and Crawford, Vincent P. (2002), "Cognition and behavior in two-person guessing games: An experimental study." Working paper. http://weber.ucsd.edu/∼vcrawfor/.

Crawford, Vincent P. (2003), "Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions." *American Economic Review*, 93: 133–49.

El-Gamal, Mahmoud and Grether, David (1995), "Are people Bayesian? Uncovering behavioral strategies." *Journal of the American Statistical Association*, 90: 1137–45.

Ericsson, K. A. and Simon, H. A. (1980), “Verbal reports as data.” 37: 215–51.

Fehr, Ernst and Gachter, Simon (2000), “Fairness and retaliation: The economics of reciprocity.” *Journal of Economic Perspectives*, 14(Summer): 159–81.

—— and Schmidt, Klaus M. (1999), “A theory of fairness, competition and cooperation.” *Quarterly Journal of Economics*, 117(3): 817–68.

Frederick, Shane, Loewenstein, George, and O’Donoghue, Ted (2002), “Time discounting and time preference: A critical review.” *Journal of Economic Literature*, 40(2): 351–401.

Gabaix, Xavier and Laibson, David (2000), “A new challenge for economics: The “frame problem.” In I. Brocas and J. Carillo (eds.), *Collected Essays in Psychology and Economics* (Oxford University Press).

Au: Please provide place of publication

Goeree, Jacob K. and Holt, Charles A. (2001), “Ten little treasures of game theory and ten intuitive contradictions.” *American Economic Review*, 91: 1402–22.

Harless, D. and Camerer, C. F. (1994), “The predictive utility of generalized expected utility theories.” *Econometrica*, 62: 1251–90.

Harless, David W. and Camerer, Colin F. (1995), “An error rate analysis of experimental data testing Nash refinements.” *European Economic Review*, 39: 649–60.

Jacoby, J., Jaccard, J., Kuss, A., Troutman, T., and Mazursk (1985), *New Directions in Behavioral Process Research* (NY: New York University).

Johnson, E. J. and Schkade, D. A. (1989), “Bias in utility assessments: Further evidence and explanations.” *Management Science*, 35: 406–24.

—— Camerer, C. F., Sen, S., and Rymon, T. (2002), “Detecting failures of backward induction: Monitoring information search in sequential bargaining.” *Journal of Economic Theory*, 104: 16–47.

—— Payne, J. W., Schkade, D. A., and Bettman, J. R. (1991), Monitoring information processing and decisions: The MouseLab system Version 5.0.

Au: Please provide Journal name

Just, M. A. and Carpenter, P. A. (1976), “Eye fixations and cognitive processes.” 8: 441–80.

Kohlberg, Elon and Mertens, Jean-Francois (1986), “On the strategic stability of equilibrium.” *Econometrica*, 54: 1003–37.

Loewenstein, G., Thompson, L., and Bazerman, M. (1989), “Social utility and decision making in interpersonal contexts.” *Journal of Personality & Social Psychology*, 57: 426–41.

Lohse, G. L. and Johnson, E. J. (1996), “A comparison of two process tracing methods for choice tasks.” *Organizational Behavior and Human Decision Processes*, 68(1): 28–43.

McKelvey, R. D. and Palfrey, T. R. (1995), “Quantal response equilibria for normal form games.” *Games and Economic Behavior*, 7: 6–38.

—— and —— (1998), “Quantal response equilibria for extensive form games.” *Experimental Economics*, 1: 9–41.

Nagel, Rosemarie (1999), “A survey of research on experimental beauty-contest games.” In D. Budescu, I. Erev, and R. Zwick (eds.), *Games and Human Behavior: Essays in Honor of Amnon Rapoport* (Mahwah, NJ: Lawrence Erlbaum).

Nisbett, Richard C. and Wilson, Timothy D. (1977), “Telling more than we can know: Verbal reports on mental processes.” *Psychological Review*, 84: 231–59.

Payne, J. W. (1976), “Task complexity and contingent processing in decision making: An information search and protocol analysis.” 16: 366–87.

—— and Braunstein, M. L. (1978), “Risky choice: An examination of information acquisition behavior.” 6: 554–61. Au: Please provide journal name

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993), *The Adaptive Decision Maker* (New York: Cambridge University Press).

Rabin, M. (1993), "Incorporating fairness into game theory and economics." *American Economic Review*, 83(5): 1281–302.

—— (2000), "Risk aversion and expected-utility theory: A calibration theorem." *Econometrica*, 68: 1281–92.

Reiskamp, Jorg and Hoffrage, U. (2000), "When do people use simple heuristics and how can we tell?" In G. Gigerenzer and P. Todd, ABC Research Group (eds.), *Simple Heuristics that Make Us Smart* (New York: Oxford University Press).

Au: Please check Rubinstein 1999 & Russo 1978 are not cited in the text.

Rubinstein, Ariel (1982), "Perfect equilibrium in a bargaining model." *Econometrica*, 50: 97–109.

—— (1999), "Experience from a course in game theory: Pre- and postclass problem sets as a didactic device." *Games and Economic Behavior*, 28: 155–70.

Russo, J. E. (1978), *Eye Fixations can Save the World: A Critical Evaluation and a Comparison Between Eye Fixations and Other Information Processing Methodologies*, J. K. Hunt, Trans. Vol. 5 (Ann Arbor, MI: Association for Consumer Research).

—— and Rosen, L. D. (1974), "An eye fixation analysis of multialternative choice." *Memory and Cognition*, 3: 267–76.

Schkade, D. A. and Johnson, E. J. (1989), "Cognitive processes in preference reversals." *Organizational Behavior & Human Decision Processes*, 44(2): 203–31.

Schotter, Andrew, Weigelt, Keith, and Wilson, Charles (1994), "A laboratory investigation of multiperson rationality and presentation effects." *Games and Economic Behavior*, 6: 445–68.

Stahl, D. and Wilson, P. (1995), "On player's models of other players: Theory and experimental evidence." *Games and Economic Behavior*, 7: 218–54.

Ståhl, Ingolf (1972), *Bargaining Theory* (Stockholm: Economic Research Institute).

Tversky, A. and Kahneman, D. (1991), "Loss aversion in riskless choice—a reference—dependent model." *Quarterly Journal of Economics*, 106(4): 1039–61.

Van Damme, Eric (1989), "Stable equilibria and forward induction." *Journal of Economic Theory*, 48: 476–96.

Weizsacker, Georg. "Ignoring the rationality of others: Evidence from experimental normal-form games." *Games and Economic Behavior*, in press. Au: Please update.