INTRODUCTION

Embodied brain activity leads to emergent computations that determine individual decisions. In turn, individual decisions, in the form of messages sent to an institution, lead to emergent computations that determine group-level outcomes. Computations can be understood in terms of a set of transformation rules, the encoding of information, and the initial conditions that together produce the computation, and we will refer to these three elements together as a computational mechanism or simply a mechanism. Neuroeconomics is interested in understanding the interrelationship between those mechanisms that exist in our evolved brains and those that exist in our constructed institutions, and their joint computation.

Game theory provides a nice middle ground for neuroeconomics studies, because it links individual decision-making to group-level outcomes with a clearly defined mechanism. The mechanism is the game tree, which specifies who gets to move when, what moves they can make, what information they have when they make their move, and how moves of different players interact to determine a joint outcome over which the players have varied interests. Non-cooperative game theory has played an important role in economic thinking, starting with the studies of imperfect competition in the late 1800s, but it was...
the publication of von Neumann and Morgenstern’s (1947) book, followed shortly by John Nash’s (1950) formulation of and proof of existence of Nash equilibrium that gave game theory its modern form. In 1994, the Nobel Prize in Economic Sciences was awarded to John Harsanyi (1920–2000), John Nash (1928–), and Reinhard Selten (1930–).

As game theory has grown in popularity, many books have become available to the reader. In addition to our review below, and the reference therein to original works, an accessible treatise is Osborne (2004).

### Extensive Form Games

A game involves two or more players. Figure 5.1 depicts two-person games. Figure 5.1(a) is an example of an extensive form game with perfect information (Kuhn, 1950). The game consists of nodes and branches that connect the nodes, the game tree. The nodes \( n_1 - n_4 \) are called decision nodes, as they each have branches connecting them to later nodes, and the nodes \( t_1 - t_5 \) are called terminal nodes. Each terminal node has a payoff vector associated with it where the top number is decision-maker 1’s payoff and the bottom number is decision-maker 2’s payoff. For convenience the branches have been labeled, L, R, LL, LR, LRL, etc. To the top left of each decision node is a number, 1 or 2, indicating that decision-maker owns, or equivalently gets to move at, that node. Decision-maker 1 owns \( n_1 \) and \( n_2 \).

**Pure Strategy Nash Equilibrium**

A play of the game is a connected path through the game tree that starts at \( n_1 \) and ends at a terminal node. Thus, (L, LR, LRL) is a play that ends at \( t_5 \). A pure strategy for a player is a choice of branch for each decision-maker at each decision node that he owns. For decision-maker 1, let \( X = \{ (L, LRL), (L, LRR), (R, LL), (R, LRR) \} \) denote the set of all possible pure strategies, and let \( x \in X \) be a particular pure strategy. Similarly, for decision-maker 2, let \( Y = \{ (LL, LRRL), (LL, LRRR), (LR, LRLR), (LR, LRRR) \} \) denote the set of all pure strategies, and let \( y \in Y \) be a particular strategy. Each strategy pair \((x, y)\) determines a play through the game tree. Thus \( x' = (L, LRL) \) and \( y' = (LR, LRR) \) determine the play (L, LR, LRL), as does the strategy pair \((x', y') = (L, LRL)\) and \( y' = (LR, LRR) \). The payoffs for decision-makers 1 and 2 can be denoted \( P(x, y) \) and \( Q(x', y') \), respectively. For example, \( P(x', y') = 30 \) and \( Q(x', y') = 60 \).

A Nash Equilibrium of a game is a strategy pair \((x^*, y^*)\) such that the following conditions hold:

\[
P(x^*, y^*) \succeq P(x, y^*) \text{ for all } x \in X
\]

\[
Q(x^*, y^*) \succeq Q(x^*, y) \text{ for all } y \in Y
\]

From the definition, it is clear that a candidate strategy pair \((x^*, y^*)\) for a Nash Equilibrium can be rejected if we can find a strategy for either player that leads to a better outcome for that player given the
other player’s strategy – i.e., if either inequality below is true:

\[ P(x', y') > P(x, y) \text{ for some } x \in X, \text{ or} \]
\[ Q(x', y') \geq Q(x, y') \text{ for some } y \in Y. \]

Thus a Nash Equilibrium strategy pair is a pair that cannot be rejected. If the inequalities in equations (5.1) and (5.2) are replaced with strict inequality signs, then we call the pair \((x', y')\) a \textit{Strict Nash Equilibrium}.

For example, \(x^* = (L, LRR)\) and \(y^* = (LL, LRRL)\) is a Nash Equilibrium in our game above. For a more general game with \(m\) players, Nash Equilibrium (see Nash, 1950) is defined as above only with \(m\) simultaneous inequalities. On the other hand, \(x^* = (R, LRR)\) and \(y^* = (LL, LRRL)\) is not a Nash Equilibrium of the game since \(P(x^*, y') > P(x', y')\).

A number of attempts have been made to write software that can calculate all of the Nash Equilibria of a game. One such example is Gambit, co-authored by Richard McKelvey, Andrew McLennan, and Theodore Turocy (2007), which can be downloaded at http://gambit.sourceforge.net. Sample Gambit output for the game above is shown in Figure 5.2a. Gambit found six Nash Equilibria, including three that involve \textit{mixed strategies} (described later). The fact that a game can have more than one Nash Equilibrium has inspired many attempts to define refinements of the Nash Equilibrium.

**Subgame Perfect Equilibrium**

One important refinement of Nash Equilibrium, due to Reinhard Selten (1975), is known as the subgame

---

**FIGURE 5.2** Solving the game using Gambit: (a) all Nash Equilibria of the game; (b) subgame perfect Nash Equilibrium of the game. Gambit is a software tool that allows you, with care, to enumerate all of the Nash Equilibria of a finite game. In Figure 5.2(a), we see the game in Figure 5.1(a) depicted in Gambit. Below the tree we see all of the Nash Equilibria of the game including mixed strategies. Gambit also has features that allow us to solve for the subgame perfect equilibrium. This is shown in Figure 5.2(b), which shows that the strategy \((R, LRL)\) and the strategy \((LL, LRRL)\) is a subgame perfect NE. Note some of the branch labels have been changed. Gambit also allows us to solve for the quantal response equilibrium (QRE) of the game described later in this chapter.
perfect equilibrium, or simply SPE, of an extensive form game. A feature of an SPE is that players have no incentive to deviate from it during the course of the game. That is, an SPE is a Nash Equilibrium strategy no matter where in the game the player begins.

To see this more clearly, consider our example game in Figure 5.1a. Note that each of the decision nodes \( n_2 \sim n_4 \) describes a subgame of the original game by simply ignoring what went before the particular node. Treat each of these nodes as a starting node of the new subgame. A strategy profile \((x^*, y^*)\) is a SPE if the relevant parts of the profile are also a Nash Equilibrium for each subgame of the original game. So, for example, while \( x = (L, LRR) \) and \( y = (LL, LRL) \) is a Nash Equilibrium of the game, it is not an SPE since \([*^*, LRR]) \neq \text{Nash Equilibrium for the subgame starting at } n_3 - i.e., decision-maker 1 would strictly prefer to play LRL. If we define the length of a subgame as the longest path to a terminal node, then we can find all of the subgame perfect equilibria of a game by working backwards through the game tree and finding all of the Nash Equilibria of the subgames, starting with subgames of the shortest length, the next shortest length, etc. For our example, \( y^* = (*, LRL) \) is the Nash Equilibrium of the subgame starting \( n_4 \), \( x^* = (*, LRL) \); \( y^* = (LL, LRL) \) is the Nash Equilibrium of the subgame of the game starting at \( n_3 \); and, finally, \( x^* = (R, LRL), y^* = (LL, LRLR) \) is a Nash Equilibrium of the game starting at \( n_1 \). These calculations are also shown in the Gambit output illustrated in Figure 5.2b. Kuhn (1953) proved that every finite extensive form game with perfect information has a SPE.

### Mixed Strategy Equilibrium

A difficulty with our definition of Nash Equilibrium described above is that not every game has such a Nash Equilibrium. The difficulty can be seen in the Rock–Scissors–Paper example shown in Figure 5.3a. In this game, both decision-makers must simultaneously choose Rock (R), Scissors (S) or Paper (P). When their choices are revealed, their payoffs are as follows: If they both choose the same, then they each get zero. If R and P are played, then P wins and the loser must pay the winner $1. If S and P are played, then S wins and the loser must pay the winner $1. Finally, if R and S

### Normal or Strategic Form Games

The extensive form game shown in Figure 5.1a has an equivalent strategic or normal form, as shown in Figure 5.1b. In a strategic form game, each player has to make a choice from a set of choices. Player DM1 chooses one of the four columns, where each column represents a pure strategy choice. Simultaneously, player DM2 chooses one of four rows corresponding to one of DM2’s pure strategies. Players’ choices together select one of the 16 cells in the \(4 \times 4\) matrix game. The cell selected is the outcome from playing the game, and each cell contains the payoff to the players if that cell is selected.

The set of Nash Equilibria for the normal form of this game is exactly identical to that described above for this game’s extensive form. Indeed, the set of Nash Equilibria for any given strategic form game is the same as the set of Nash Equilibria for that same game expressed in extensive form. The reason is that Nash Equilibrium is defined in terms of available strategies. It does not matter to Nash Equilibrium analysis when those strategies are executed, or how they are described.
are played, then R wins and the loser must pay the winner $1. To see that there is no Nash Equilibrium as defined above, note that if DM1 plays R then DM2 strictly prefers P; but if DM2 plays P then DM1 strictly prefers S, and so on, resulting in an endless cycle, or equivalently no solution to the inequalities (5.1) and (5.2) above. However, we can find a Nash Equilibrium of this game if we allow DM1 and DM2 to choose mixed strategies as defined below.

We can denote a mixed strategy for decision-makers 1 and 2 as probability distributions, p and q, over X and Y respectively. Thus, for any given Morgenstern preferences, we can assume the players’ preferences regarding strategies are ordered according to expected payoffs (Bernoulli, 1738; von Neumann and Morgenstern, 1944; see also Chapter 3 of this volume). Thus, for any given p and q, DM1 and DM2’s respective payoffs are given by:

\[
EP(p, q) = \sum_{x \in X} \sum_{y \in Y} p(x)q(y)E(x, y)
\]

A pure strategy \((x, y)\) is now a special case of a mixed strategy where \(p(x) = 1\) and \(q(y) = 1\). A mixed strategy Nash Equilibrium is a \(p^*, q^*\) such that \(EP(p^*, q^*) \geq EP(p, q^*)\) for all mixed strategies \(p\) and \(q \geq EP(p^*, q)\) for all mixed strategies \(q\). For the Rock–Scissors–Paper game, there is one Nash Equilibrium in mixed strategies \(p = (1/3, 1/3, 1/3)\) and \(q = (1/3, 1/3, 1/3)\).

More generally, if we have a strategic form game with \(n\) players, indexed by \(i\), each of whom have von Neumann-Morgenstern preferences, then we can define the all the remaining strategies of the \(n-1\) players as \(p_{-i} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n)\) and payoffs can be defined by \(EU(p_{-i})\). We can now extend our definition of Nash Equilibrium to mixed strategies as follows: the mixed strategy \(p^*\) is a Nash Equilibrium if, and only if, for each player \(i\),

\[
EU_i(p^*, p_{-i}) = \max_{p_i} EU_i(p_i, p_{-i})
\]

We can identify these players von Neumann-Morgenstern payoff function as

\[
EU_i(p) = \sum_{x \in X} p_i(x)E_i(x, p_{-i})
\]

where \(i\)'s pure strategy replaces the mixture.

Thus every mixed strategy Nash Equilibrium has the property that each player’s expected equilibrium payoff is the player’s expected payoff to any of the actions used with positive probability. For our example of Rock–Scissors–Paper, given the Nash Equilibrium strategy of playing each strategy with 1/3 probability, \(EP(p^*, q^*) = 0\) for \(x = \) Rock, Scissors, or Paper, and \(EQ(p^*, q) = 0\) for \(y = \) Rock, Scissors, Paper, verifying that \(p^* = q^* = (1/3, 1/3, 1/3)\) is a Nash Equilibrium.

Nash (1950) demonstrated that every strategic game with a finite number of players with von Neumann-Morgenstern preferences, and a finite number of strategies, has a Nash Equilibrium.

If we modify the Rock–Scissors–Paper game by forbidding DM2 to play Paper, then we have the game depicted in Figure 5.3b. The reader may want to verify that the Nash Equilibrium of this game is \(p^* = (2/3, 0, 1/3)\) and \(q^* = (2/3, 1/3)\).

### Games with Incomplete Information

John Harsanyi (1967/68) formalized the idea of a game with incomplete information. Consider the standard trust game shown in Figure 5.4a. An interpretation of this game is that player 1 chooses between splitting 20 equally with player 2, or sending the 20 to player 2 in which case it becomes 40. Player 2 can then either “cooperate” by choosing to give 15 to player 1 and 25 to herself, or “defect” by keeping the entire 40. The subgame perfect Nash Equilibrium is for player 1 to choose R at \(n_1\) and player 2 to choose R at \(n_2\). That is, if an earnings-maximizing player 2 is given the chance to move, he should “defect” by choosing R. Player 1 anticipates this, and thus chooses not to give player 2 a move, and instead divides the 20 equally.

It is possible that humans have evolved a propensity to cooperate in games like this. For example, people may feel guilty when, as player 2, they defect with the move R. Suppose, furthermore, that guilt reduces a player’s final payoff by some amount. A low-guilt person \(G_L\) may only experience a small payoff loss (say 5), while a high-guilt person \(G_H\) may experience a much higher payoff loss (say 20). Finally, we will assume that player 2 knows whether he is a low-guilt or high-guilt type person, but player 1 does not know player 2’s type. We can depict this game as a Bayesian trust game with incomplete information, as shown in Figure 5.4b.

There are two important additions to the Bayesian trust game. First, there is a starting move by nature at node \(n_0\) that determines the type of player 2. Instead of providing a label to the branches that Nature can choose, we indicate the probability that Nature will choose the branch. So with probability \(1/4\) Nature will move left and the trust game will be played with
DM2s that DM1 encounters feel enough guilt (G H) to modify mental results is to use a Bayesian Game, Figure 5.4(b), where some more than 50% of DM2s reciprocate. One way to explain the experiment is shown by the dotted line between n2 and n4, indicating that DM1 always trusts player 2, as shown by the arrows resulting in an expected payoff of 11.25 for player 1. The break-even point for player 1 is a belief that at least 2/3 of the population will feel strong guilt regarding cheating. Thus we can see how optimistic player 1’s will try to cooperate by playing L, while more pessimistic player 1’s will play R.

Information sets act as a general formalism for denoting how much information a player has about the previous moves in an extensive form game. Games with perfect information are a special case where all information sets contain only one node, i.e. all players know perfectly what path of the game they are on when it is their turn to make a move. Games with at least one information set containing more than one node are called games with imperfect information.

When a player is at an information set, he must act as though he has only one move – that is, the same move at each node in the set – but realize that his move may be along different paths of the game.

A typical game of imperfect information (with many variations) is the simple card game shown in Figure 5.5a. Two players ante up by each putting a dollar in the pot. Nature deals one card to player 1. There is a 50–50 chance that the card is high or low. A high card, if shown, wins the pot; a low card, if shown, loses the pot. A fold also loses the pot. At this point, nodes n1 and n2, player 1 sees the card and can decide to show (and win a dollar if high, or lose a dollar if low) or raise the bet by adding a dollar to the pot. Now it is player 2’s turn to move without knowing player 1’s card; thus the information set containing n1 and n2. Player 2 can fold (thus losing a dollar) or call by adding a dollar to the pot. A call forces player 1 to show the card, and determines the winner.

A game with imperfect information has perfect recall if every player remembers whatever he knew in the past. We can now distinguish a mixed strategy for an extensive form game as a probability mixture over pure strategies from a behavioral strategy where
The players pick a probability measure at each information set in the game with the property that each distribution is independent of every other distribution. An important theorem is that any for any mixed strategy of a finite extensive form game with perfect recall, there is an outcome-equivalent behavioral strategy. Note, a behavioral strategy and a mixed strategy are outcome-equivalent if for every collection of pure strategies of the other players the two strategies induce the same outcome. An immediate consequence is the equivalence of Nash Equilibrium for games with perfect recall, there is an outcome-equivalent behavioral strategy. Any important theorem is that any for any mixed strategy of a finite extensive form game with perfect recall, there is an outcome-equivalent behavioral strategy. Note, a behavioral strategy and a mixed strategy are outcome-equivalent if for every collection of pure strategies of the other players the two strategies induce the same outcome. An immediate consequence is the equivalence of Nash Equilibrium for games with perfect recall, there is an outcome-equivalent behavioral strategy.
where $\epsilon_1 = (\epsilon_{i1}, \ldots, \epsilon_{im})$ is distributed according to the distribution function $f_i(\epsilon)$ such that the expected value of $\epsilon_i$ is 0. Given $u'(p)$ and $f$, player $i$'s best response is to choose $s_i$ such that $u'_i(p) \equiv u'_i(p^*)$ for $k = 1, \ldots, m$. Note that this relaxes the assumption that players make decisions without errors. In a quantal response equilibrium, best response functions take into account that the game’s players make decisions with error: there is some chance that they won’t implement a strategy successfully.

Now, given each player’s best response function and distribution $f_i$, calculate $\sigma_i^*(u_i'(p))$ as the probability that player $i$ will select strategy $j$ given $u'(p)$. Given $f$, a quantal response equilibrium for a finite strategic game is a mixed strategy $p^*$ such that for all players $i$ and strategies $j$, $p^*_i = \sigma_i^*(u_i'(p^*))$. A quantal response equilibrium exists for any such game. A standard response function is given by the logistic response function

$$\sigma_i^*(x_i) = (\exp(\lambda x_i) / \sum_{j=1}^m \exp(\lambda x_j))$$

where $x_j = u_j(p)$.

Using Gambit and the above logistic response function, we calculate the quantal response equilibrium for Selten’s Horse (Figure 5.5b) and see that as $\lambda \to \infty$, $p \to (C, c, R)$; however, McKelvey and Palfrey (1995) provide a counter-example to show that the limiting quantal response equilibrium does not always correspond to Selten’s trembling hand perfection.

It is worthwhile to note that Quantal response equilibrium can be interpreted as a reinforcement learning model (Goeree and Holt, 1999).

**GAME THEORY EXPERIMENTS**

Game theory experiments test game theory. To be interesting to neuroeconomists typically requires that the experiment goes further than this – for example, by informing the role of emotion in decision or the number of cognitive “types” that exist in a population. This might involve brain-imaging, but need not necessarily do so. In this section we describe the design, practice, and results of game theory experiments that do not include an imaging component but that are especially relevant to neuroeconomic research.

**Design and Practice**

An important feature of laboratory game theory experiments is that a participant’s decisions can be highly sensitive to the specifics of the implementation. An implication is that many game theory experiments are powerful tools for uncovering critical features of the human decision process that might be relatively difficult to detect outside of controlled environments. In addition, like the best theory, the results of game theory experiments might be useful in shedding light on behavioral principles applicable to a large number of decision environments.

The particulars of any experiment design depend on the hypotheses it intends to test. However, there are general design considerations common to any game theory experiment that will be reviewed briefly here. Readers interested in more thorough recent discussions of design and analysis considerations should consult Houser (2008), Camerer (2003, especially Appendix A1.2), Kagel and Roth (1995), Friedman and Sunder (1994), and Davis and Holt (1993). In addition, the outstanding book by Fouraker and Siegel (1963) offers an early but still relevant discussion and defense of procedures in experimental economics. Their work draws attention to instructions, randomization, anonymity, and salient rewards, each of which remains fundamental to experimental economics procedures, as discussed below.

**Instructions**

It might be expected that decisions would be rather insensitive to the nature of a game’s description, so long as the instructions were clear and complete. The reason this is not the case is that instructions not only describe but also frame an environment, and behavior is highly sensitive to framing. For example, using the word “partner” instead of “counterpart” to describe a matched participant in an experiment can affect decisions substantially (see Houser et al., 2004, for further discussion and application to the challenge this raises for interpretation of cross-cultural experiments.) As a result, it is important to make an experiment’s instructions consistent among various sessions of the same treatment. This is often facilitated by providing a written form of the instructions to subjects, and then reading it to them at the beginning of each session.

**Randomization**

The role of randomization also cannot be overstated. One reason is that it is necessary for the validity of a variety of widely-used analysis procedures – see, for example, Houser (2008) for elaboration. More generally, the appropriate use of randomization avoids confounding influences on the results. For example, as noted by Fouraker and Siegel (1963), subjects might...
differ in personality traits or preferences for money, and these differences might also be correlated with the time at which a subject arrives at the experiment. Random assignment of subjects to treatments and roles within the experiment helps to ensure that such differences do not systematically affect an experiment’s outcome.

**Anonymity**

To guarantee anonymity, participants are randomly assigned counterparts, visually separated from each other, and asked to remain silent for the duration of the experiment. By ensuring that participants do not know with whom they are matched, the possibility that decisions will be based on perceptions unrelated to the decision environment under study is largely eliminated. Random and anonymous matching also substantially mitigates the possibility of (unwanted) collusion among participants, which might otherwise confound inferences and interpretations.

It should be emphasized that randomization and anonymity effectively control for differences in individual characteristics only to the extent that each treatment in the experiment uses different participants drawn from a common “cohort,” or group with the same distribution of characteristics (e.g., demographic and personality). Often, this is ensured by using a university’s general student population to select participants. An alternative is to study the decisions of the same people in multiple treatments. While this can in some cases be efficient, it also raises special design and analysis considerations (e.g., controlling for effects due the order in which treatments are experienced, as well as the fact that repeated and perhaps correlated observations are obtained from each person).

**Salient Rewards**

A hallmark of experimental economics, “salient rewards” refer to monetary payments that vary according to a person’s decisions in an experiment. Vernon Smith formalized the importance of this procedure with the publication of his Induced Value Theory (Smith, 1976). As long as it is assumed that people prefer more money to less, applying the theory to experiments requires only that real money values are assigned to tokens earned in an experiment. Intuitively, the advantage to doing this is that it raises confidence that participants will recognize the economic incentives implied by the game environment. Induced value theory rigorously justifies experimental tests of game theory, and as such has facilitated the development of new theories incorporating “social preferences” that have been especially relevant to the development of neuroeconomics.

The use of salient rewards in economics experiments stands in sharp contrast to the use of hypothetical rewards common in the psychology literature. As a practical matter, the importance of salient rewards is an empirical question whose answer is likely sensitive to the environment under study (for example, Holt and Laury (2002) compare risk elicitation under both hypothetical and salient reward circumstances). An accepted principle dating from Smith (1965) is that, in relation to hypothetical environments, salient rewards are likely to reduce the variability of decisions among subjects.

Salient rewards are perhaps most transparent in so-called “one-shot” games in which players make a decision, receive their earnings, and then the experiment ends (for example, standard implementations of dictator and ultimatum games). Interestingly, in imaging studies it is typically necessary (for technical reasons) to modify the experiment’s protocol so that these games become “multi-shot” or “repeat-single”. This means that the game is played several times instead of once, with earnings usually determined by a random draw from one of the completed games. Participants are usually anonymously matched with different players for each game (so-called “Strangers” matching) in order to avoid effects caused by, for example, “reputation,” meaning beliefs developed about their counterpart’s likely decisions based on play during the course of the experiment.

**Experiments with Normal Form Games**

**Prisoner's Dilemma and Public Goods Games**

Prisoner’s dilemma (PD) and public goods (PG) games are used to study “social dilemmas” that arise when the welfare of a group conflicts with the narrow self-interest of each individual group member. For example, in a typical two-player PD, each player can choose either to “cooperate” or “defect.” Payoffs are symmetric, and chosen so that the sum of the payoffs is greatest when both choose “cooperate” and least when both players choose “defect.” However, each player earns the most if he chooses to “defect” when the other cooperates. Thus, the unique subgame perfect Nash Equilibrium of this environment is for both players to defect.

The structure of PG games is similar, but they are typically played in larger groups. In a typical PG game, each member of a group of four people is allocated $10. Group members simultaneously decide how to allocate their endowment between two
“accounts,” one private and one public. The private account returns $1 to the subject for each dollar allocated to that account. In contrast, every dollar invested in the public account doubles, but is then split equally among the four group members ($0.50 each). Thus, like the PD game, group earnings are maximized at $80 if everybody “cooperates” and contributes everything to the public account, in which case each of the four participants will earn $20. However, if three subjects contribute $10 each while the fourth “free-rides” and contributes nothing, then the “free-rider” will earn $25 (the best possible outcome for him). Like the PD game, each group member has the private incentive to contribute nothing, and the unique equilibrium perfect Nash Equilibrium occurs when each subject contributes zero to the group account.

Standard results for PD and PG games are discussed at length elsewhere (see, for example, Davis and Holt, 1993; Ledyard, 1995). The key early finding was that, in aggregate, cooperation occurs about half of the time in PD games, and that about half of the aggregate endowment is contributed to the “public” account in a PG game. It is also routinely found that aggregate cooperation decays when these games are repeated, though cooperation usually remains above zero even with a relatively large number of repetitions (say 30). Though the specific patterns of cooperation can vary with the particulars of the game, the substantive finding that people cooperate in social dilemmas is robust. Results from these early games opened the door for “psychological” game theory (Geanakoplos et al., 1989) in which concepts such as reciprocity and altruism play important roles.

PG games continue to be widely studied, and have proven a sharp guide for theories of social preferences (see Chapter 15 of this volume). One reason is that it is simple to develop designs for these games that allow compelling evidence on critical issues in social preference theory. For example, Gunthorsdottir et al. (2007) provide rigorous data from a PG experiment that show that (positive) reciprocity is more important than altruism in driving cooperation. Another reason is that PG games provide rich data on individual decision patterns. For example, PG data starkly reveal that individuals fall into cleanly described “types” (Kurzban and Houser, 2005), and stress that any theory of social preferences that does not account for individual differences is substantively incomplete.

Coordination Games

Unlike standard PD or PG games, many games have multiple equilibria that require coordination. For example, a simple two-player, two-alternative (say “A” and “B”) “matching” game will pay each player $1 if they both choose “A” or both choose “B,” but will pay each of them nothing if their choices do not match. In these environments, a key role for experiments is to help discover the relative likelihood that a particular equilibrium might be played, as well as the features of the environment (including participant characteristics) that determine this likelihood.

The large literature in coordination games cannot be discussed here, but is well reviewed by Camerer (2003: Chapter 7); this literature also suggests several “stylized facts” regarding play in these games. These include that (i) coordination failure is common; (ii) repeated play does not reliably converge to a Pareto-efficient outcome (meaning that no reallocation can make all players simultaneously better off); (iii) the nature of convergence depends on the information available to players and how the players are matched; and (iv) whether and how players are allowed to communicate can have substantial effects on outcomes. Although important challenges arise in its analysis (Houser and Xiao, 2008), communication provides perhaps the richest data for understanding decisions in social environments that require coordination.

Experiments with Extensive Form Games

Ultimatum Games

The ultimatum game, introduced by Guth et al. (1982), is a simple take-it-or-leave-it bargaining environment. In ultimatum experiments, two people are randomly and anonymously matched, one as proposer and one as responder, and told they will play a game exactly one time. The proposer is endowed with an amount of money, and suggests a division of that amount between himself and his responder. The responder observes the suggestion and then decides whether to accept or reject. If the division is accepted, then both earn the amount implied by the proposer’s suggestion. If rejected, then both the proposer and responder earn nothing.

The key result of ultimatum experiments is that most proposers offer between 40% and 50% of the endowed amount, and that this split is almost always accepted by responders. When the proposal falls to 20% of the endowment it is rejected about half of the time, and rejection rates increase as the proposal falls to 10% and lower. As discussed by Camerer (2003: Chapter 2), ultimatum game results are highly robust to a variety of natural design manipulations (e.g., repetition, stake size, degree of anonymity, and a variety of demographic variables).
An important exception to robustness is reported by Hoffman and Spitzer (1985), who show that offers become significantly smaller, and rejections significantly less frequent, when participants compete for and earn the right to propose. An explanation is that this procedure changes the perception of “fair,” and draws attention to the importance of context in personal (as compared to market) exchange environments. These effects might also stem from varying the degree of anonymity among the subjects, or between the subjects and the experimenter (Hoffman et al., 1996).

A key focus of recent ultimatum game research has been to understand why responders reject low offers. Economic theory based on self-interested preferences suggests that responders should accept any positive offer and, consequently, proposers should offer the smallest possible positive amount. We review some well-known research on the topic of responder rejections in the “Neuroeconomics experiments” section below.

Trust Games

Joyce Berg, John Dickhaut and Kevin McCabe introduced the popular trust game in 1995. Two participants are randomly and anonymously matched, one as investor and one as trustee, and play a one-shot game. Both participants are endowed with $10. The investor can send some, all, or none of his $10 to the trustee. Every dollar sent by the investor is tripled. The trustee observes the (tripled) amount sent, and can send some, all, or none of the tripled amount back to the investor. The amount sent by the investor is a measure of trust; the amount returned by the trustee is a measure of trustworthiness.

Berg et al. (1995) reported that investors send about 50% of the endowment on average, and trustees generally return the amount sent. There is more variance in amounts returned than in amounts sent. Indeed, Berg et al. (1995) reported that fully 50% of trustees returned $1 or less. Camerer (1993: Chapter 2) described a variety of studies that replicate and extend these first results. As we discuss further below, this game is also widely used in neuroeconomics experiments, including (i) purely “behavioral” experiments with healthy volunteers that provide evidence on the role of, for example, emotion on decision; (ii) “lesion” studies that examine the behavioral consequences of brain damage (or temporary disruption with transcranial magnetic stimulation (TMS)); (iii) examinations of drug effects on economic decisions; (iv) skull-surface based measurement of brain electrical activity during decision tasks using electroencephalography (EEG) or magnetoencephalography (MEG); and (v) real-time whole brain imaging using functional magnetic resonance imaging (fMRI) during an economic decision task. A comprehensive review of the leading procedures to draw inferences from brain data can be found in Toga and Mazziotta (2002).

Although each method has unique advantages, over the past decade fMRI has emerged as the dominant technique. The reason is that it is a relatively easily implemented, non-invasive procedure for scientific inference with respect to real-time brain function in healthy volunteers during decision tasks. It is therefore worthwhile to comment briefly on the design and practice of fMRI experiments. Much more detailed discussion can be found in any of a number of recent textbooks that focus exclusively on this topic (Hüttel et al., 2004 is an excellent and especially accessible source).

Overview

An fMRI neuroeconomics experiment correlates brain activity with economic decision-making. However, it does not directly measure neural activity. Rather, evidence on cerebral blood flow is obtained, which Roy and Sherrington (1890) discovered is correlated with underlying neuronal activations. The reason is that active neurons consume oxygen in the blood, leading the surrounding capillary bed to dilate and (with some delay) to an increase in the level of oxygenated blood in the area of neural activity. It turns out that this “hemodynamic response” can be detected and traced over time and (brain) space. Although fMRI technology is rapidly improving, most early studies reported data with temporal resolution of 1 or 2 seconds, with each signal representing a three-dimensional rectangular “voxel” measuring 2 or 3 millimeters on each side and containing literally millions of neurons.

Design

The design of an fMRI neuroeconomics experiment should ensure that the hemodynamic, or blood-oxygen level dependent (BOLD), response can be...
detected, as well as reliably traced to neural activity associated with the decision processes of interest. A technical constraint in this regard is that the BOLD signal is quite weak, with typical responses being just a few percentage points from baseline measurements made by a typical scanner. An important implication is that neuroeconomic experiments typically require multiple plays of the same game and an averaging of the signals produced therein. That is, single-shot studies are not possible with current technology, and the design strategy must take this into account. A second implication of the weak signal is that other sources of signal variation, such as motion of the subject in the scanner, must be strictly controlled at data collection, and again accounted for during data "preprocessing."

Analysis
The analysis of fMRI data occurs in two stages. The first stage is "preprocessing," the components of which include (i) image realignment to mitigate variation in the data due to head motion; (ii) image standardization to facilitate comparisons among brains of different participants; and (iii) image smoothing to reduce high-frequency voxel specific noise. How different preprocessing approaches affect second-stage inference is the subject of active research (see, for example, Chen and Houser, 2008).

The second stage involves analyzing the (preprocessed) data and drawing inferences about activation patterns. Regardless of the approach used to do this, it is necessary to confront the issue that imaging data has a massive spatial-panel structure: the data include observations from thousands of spatially and temporally characterized voxels. The analysis strategy should allow for the possibility that proximate voxels might have a correlated signal structure, especially because appropriate inference requires accounting for multiple comparisons (see Tukey, 1991, for an accessible discussion of this issue).

Neuroeconomics Experiments with the Trust Game
The Berg et al. (1995) trust game (or close variants) has been conducted thousands of times and has played an important role in shaping economists’ view of trust and reciprocity. The trust game has also proved a useful paradigm in neuroeconomics. Indeed, it was used by McCabe and colleagues in their 2001 fMRI study of cooperation, which also turns out to be the first published imaging study of economic exchange.

McCabe et al. (2001) reasoned that cooperative economic exchange requires a theory-of-mind (ToM). They thus hypothesized that medial prefrontal cortex, which had been previously implicated in ToM processing (Baron-Cohen, 1995), would also mediate cooperative economic exchange. To test this hypothesis, they asked subjects in a scanner to play variants of a trust game multiple times either with human counterparts outside the scanner or with a computer counterpart. All trust games were "binary" (of the form described by Figure 5.4a), in the sense that both the investor and trustee chose from one of two alternatives, either "cooperate" or "defect." The computer played a known stochastic strategy, and scanner participants were informed prior to each game whether their counterpart was a human or a computer.

Of the study’s twelve subjects, seven were found to be consistently cooperative. Among this group, medial prefrontal regions were found to be more active when subjects were playing a human than when they were playing a computer. On the other hand, within the group of five non-cooperators there were no significant differences in prefrontal activation between the human and computer conditions. It is interesting to note that ToM imaging studies caught on quickly, and that the areas identified by McCabe et al. (2001) have also been found by others (see Chapter 17 of this volume for a review of the ToM neuroeconomics literature).

Another important imaging (positron emission tomography) experiment with a trust game was reported by de Quervain and colleagues (2004; see also Chapter 15 of this volume). This study sought to provide evidence on the neural basis of punishment, and in particular to investigate whether brain areas including the dorsal striatum are activated when punishing another who has abused their trust. To assess this, the authors had two anonymous human players, A and B, make decisions in a binary trust game. Both players started with 10 MUs (monetary units), and player A could either trust by sending all 10 MUs to B, or send nothing. If A chose to trust, then the 10 MUs were quadrupled to 40, so that B had 50 MUs and A had zero MUs. B could then send 25 MUs to A, or send nothing and keep the entire 50 MUs. Finally, following B’s decision, A could choose to punish B by assigning up to 20 punishment points. In the baseline treatment, each point assigned reduced A’s earnings by 1 MU and B’s earnings by 2 MUs.

This game was played in a variety of conditions, in order to ensure that the appropriate contrasts were available to assess punishment effects. In addition to the baseline, key treatment variations included the following: (i) a random device determined B’s
back-transfer, and punishment worked as in the baseline; (ii) B determined the back-transfer, but punishment points were free for A and removed 2 MUs from B’s earnings; (iii) B determined the back-transfer, and punishment points were only symbolic in the sense that they were free for A and they also did not have earnings implications for B. With these contrasts in hand, the authors were able to draw the inference that effective (but not symbolic) punishment is associated with reward, in the sense that it activates the dorsal striatum. Moreover, they found that subjects with stronger activations in that area were more likely to incur greater costs in order to punish.

Recently, Krueger et al. (2007) have found evidence for two different mechanisms for trust in repeated, alternating-role trust games with the same partner. One system for trust uses anterior paracingulate cortex in early trials, which is extinguished in later trials and replaced by activation in the septal region of the brain. Bold activations in these areas are interpreted as characterizing a system of unconditional trust in the person. Another system shows no early activation in anterior paracingulate cortex but does show a late activation consistent with the behavioral responses of subjects to be less trustworthy when temptation is greatest. This is interpreted as characterizing a system of conditional trust, as first movers learn to avoid trusting their partner when temptations to defect are high.

A large number of other trust games have been studied with various motivations. In this volume, trust games play a role in the designs discussed in Chapters XXX.

Neuroeconomics Experiments with the Ultimatum Game

Neuroeconomics experiments with the ultimatum game have been conducted with the primary goal of shedding light on reasons for rejections of unfair offers. Because a person earns a positive amount by accepting the offer, and earns nothing by rejecting, the decision to reject offers has puzzled economists. We here review three innovative studies on this topic, each of which uses a different method: a behavioral study by Xiao and Houser (2005), an fMRI study by Sanfey et al. (2003), and rTMS results reported by Knoch et al. (2006).

Xiao and Houser (2005) studied the role of emotion expression in costly punishment decisions. A substantial literature suggests humans prefer to express emotions when they are aroused (see, for example, Marshall, 1972). The results obtained by Xiao and Houser (2005) suggest that the desire to express negative emotions can itself be an important motivation underlying costly punishment.

In ultimatum game experiments conducted by Xiao and Houser (2005), responders have an opportunity to write a message to their proposer simultaneously with their decision to accept or reject the proposer’s offer. Xiao and Houser found that, compared with standard ultimatum games where the only action responders can take is to accept or reject, responders are significantly less likely to reject the unfair offer when they can write a message to the proposers. In particular, proposers’ offers of $4 (20% of the total surplus) or less are rejected 60% of the time in standard ultimatum games. When responders can express emotions, only 32% reject unfair offers, and this difference is statistically significant.

The messages written in Xiao and Houser’s (2005) emotion expression game were evaluated using a message classification game with performance-based rewards (for discussion, see Houser and Xiao, 2007, and also the related “ESP” game of von Ahn, 2005). Evaluators were kept blind to the research hypotheses as well as decisions made by participants in the emotion expression game. The vast majority of those who accepted offers of 20% or less wrote messages, and all but one of those messages were classified as expressing negative emotions. An interpretation is that costly punishment decisions occur in part as a way to express dissatisfaction. Earnings maximizing decision-making, therefore, is promoted when less expensive channels are available for the purpose of emotion expression.

Sanfey et al. (2003; see also Chapter 6 of this volume) is an early fMRI study of the ultimatum game. In this study, participant responders faced either confederate proposers or computers, so that each responder faced exactly the same set of fair (equal split) and unfair offers (between 70% and 90% to the proposer). The brain images revealed that, in comparison to fair offers from human or any computer offers, when the responders were faced with unfair offers from humans there was greater activation in the bilateral anterior insula, the anterior cingulate cortex (ACC), and the dorsolateral prefrontal cortex (DLPFC). The computer condition provides the contrast necessary to rule out the possibility that the source of the activation is the amount of money, thus providing evidence that activations are due to the “unfair” intentions of humans. Moreover, Sanfey et al. found that activation in the insula correlated positively with the propensity to reject unfair offers. Because the insula has been implicated in the processing of unpleasant emotions (Calder et al., 2001), this
result is convergent evidence that negative emotions underlie the rejection decision in the ultimatum game. The complexities of the neural networks underlying rejection decisions are underscored by results reported by Knorr et al. (2006). These researchers used repetitive transcranial magnetic stimulation (rTMS) in order to disrupt the left or right DLPFC. They found that the rate of rejection of maximally unfair offers (20% was the least amount that could be offered) was just 10% when the right DLPFC was disrupted. On the other hand, the rejection rate of unfair offers was equal to the baseline, 50%, when the disruption was to the left DLPFC. The authors concluded that right, but not left, DLPFC plays an important role in overriding self-interested impulses, which adds another piece to the puzzle that is the neural underpinning of costly punishment decisions.

Other ultimatum game studies are reviewed in various chapters in this volume, as follows.

Towards a Neuroeconomic Theory of Behavior in Games

Cognitive neuroscience has made great progress regarding the neural basis of perceptual decision-making (see, for example, Gold and Shadlen, 2007), as well as value-based decision-making (Glimcher et al., 2005). Models of decision-making based largely on single cell firing in monkeys assumes that neurons encode a sequential probability ratio test (Wald and Wolfowitz, 1947), to decide statistically among competing hypotheses. Within this framework mixed strategies can be explained at the level of neuronal noise (Glimcher et al., 2005; Hayden and Platt, 2007), although how noise biases probabilities toward optimal strategies is less understood. It is even less clear how these models of decision-making should be extended to games involving other persons.

When individuals evaluate a game tree, they make choices which they expect will result in a desired payoff goal. One approach to solving this problem is to rely on reinforcement learning (Sutton and Barto, 1998) alone, as calculated by the QRE of the game. Such an approach is parsimonious, and would involve only the goal-directed learning parts of the brain (that is, the ventral and dorsal striatum) together with a method for encoding strategies (most likely in the prefrontal cortex) and their payoff equivalents (for example, in pre-motor regions of the brain and the lateral intraparietal area or other parietal areas encoding expected utility maps) (Montague et al., 2006). However, one problem with this approach is the relatively long length of time it would take people to learn the QRE of the game. Thus, necessary additions to a reinforcement learning theory of game-playing would be various mechanisms for sharing mental states that would improve the brain choice of an initial strategy and allow the brain to weigh information appropriately and update goals in order to learn more quickly its best strategic choices (starting points for such models might include, for example, Camerer and Ho, 1999, or Erev and Roth, 1998).

Initial strategies are likely to be chosen based on an examination of payoffs leading to a goal set, where a goal set should be understood as the set of all potentially desired outcomes. One unknown is how large a goal set the brain will try to handle. For example, in the game shown in Figure 5.1a, player 1 will see $t_1$ with a payoff of 50 and the payoff of 40 at $t_2$ as his goal set from an initial set of payoffs of {50, 40, 30, 20, 0}. In the game shown in Figure 5.4a, player 1 may choose (15, 10) as his goal set from the set of possible payoffs {15, 10, 0}. How players choose their goal sets and edit them over time then becomes a critical feature of such a model. For example, are people more likely to include high payoff outcomes in their initial goal sets?

Given a goals set, a player must identify the paths that will lead to his desired goals. Since each terminal node is isomorphic to a path in the tree, there is a 1–1 and invertible function $f$ which maps the set of goal sets $G$ into the set of game paths $P$, and therefore there is set of decision nodes that are “critical” to a player’s goals in that at a critical node paths diverge. For example, in Figures 5.1a and 5.4a, a critical node for player 1 is $n_1$. Since it is at critical nodes that players make commitments to a proper subset of their goal sets, we expect the brain to weigh the evidence for each path using some form of forward induction and choose based on the resulting accumulation of support for a given strategy.

The next step is to assess who else owns decision rights along the path towards $t_j$ and what their incentives might be. So, for example, in Figure 5.1 player 2 controls the node $n_2$ and might like 60 at $t_2$ compared to 50 at $t_1$. If this possibility is likely enough, then player 1 may simply decide to play R and get 40. However, player 1 might also try mentally to simulate player 2’s mind to induce how player 2 might react at node $n_2$. Player 1 might reason that player 2 will see that there is a risk to trying for $t_2$ since player 1 controls the node $n_2$. But why would there be any risk? A simple answer is that when player 1 took the risk to try for 50, he also made an emotional commitment to punish player 2 if he tried for 60. Notice the decision to punish requires two things; an assessment of shared attention over the fact that player 1 has taken...
a risk to achieve 50, and an assessment by player 1 that player 2 can empathize with player 1’s emotional commitment to punishment.

As part of the forward induction at critical nodes, players are also likely to evaluate the person they are playing as suggested by the nature of the Bayesian trust game shown in Figure 5.4b. In this case, experiential priors from similar situations may bias the players’ beliefs (weighing) of the game they are in. When results are evaluated, they will then be updated based on reinforcement learning systems (as a slow learning process) or through much faster emotional responses, such as found in insula responses.

CONCLUSION

Neuroeconomics research helps to disentangle the complex interrelationships between the neural mechanisms with which evolution has endowed our brains, the mechanisms that our brains have built into our external institutions, and the joint computations of these mechanisms from which social and economic outcomes emerge. Game theory provides a convenient platform for neuroeconomics studies because it formally connects the strategic decisions of multiple individuals to group-level outcomes through a precisely defined mechanism.

We have seen that game theory can entail substantial abstraction. While the level of abstraction created by game theory has substantial advantages, it can also create uncertainty with respect to the way in which laboratory participants perceive the game environment. This can lead to difficulties in interpreting participants’ decisions, especially when those decisions are at odds with our constructed notions of rationality. It might be tempting to attribute the failure of a theory to participants’ failures to understand the incentives associated with a game. An alternative explanation is that the decisions of “irrational” participants are fully rational, but from an alternative perspective (e.g., “ecological rationality,” as promoted by Smith, 2007).

Neuroeconomics can help to distinguish between these explanations, and is certain to play a fundamental role in the process of discovering how people decide.

References


5. EXPERIMENTAL NEUROECONOMICS AND NON-COOPERATIVE GAMES


