Against Game Theory

Gale M. Lucas
University of Southern California

Mathew D. McCubbins
Duke University

Mark Turner
Case Western Reserve University

Abstract: People make choices. Often, the outcome depends on choices other people make. What mental steps do people go through when making such choices? Game theory, the most influential model of choice in economics and the social sciences, offers an answer, one based on games of strategy like chess and checkers: the chooser considers the choices that others will make and makes a choice that will lead to a better outcome for the chooser, given all those choices by other people. It is universally established in the social sciences that classical game theory (even when heavily modified) is bad at predicting behavior. But instead of abandoning classical game theory, those in the social sciences have mounted a rescue operation under the name of "behavioral game theory." Its main tool is to propose systematic deviations from the predictions of game theory, deviations that arise from character type, for example. Other deviations purportedly come from cognitive overload or limitations. The fundamental idea of behavioral game theory is that, if we know the deviations, then we can correct our predictions accordingly, and so get it right. There are two problems with this rescue operation, each of them fatal. (1) For a chooser, contemplating the range of possible deviations, as there are many dozens, actually makes it exponentially harder to figure out a path to an outcome. This makes the theoretical models useless for modeling human thought or human behavior in general. (2) Modeling deviations is helpful only if the deviations are *consistent*, so that scientists (and indeed decision-makers) can make predictions about future choices on the basis of past choices. But the deviations are not consistent. In general, deviations from classical models are not consistent for any individual from one task to the next or between individuals for the same task. In addition, people's beliefs are in general not consistent with their choices. Accordingly, all hope is hollow that we can construct a general behavioral game theory. What can replace it? We survey some of the emerging candidates.

Keywords: Biases, Game Theory, General Equilibrium Theory, Behavioral Game Theory, Social Dilemmas, Experimental Economics, Heuristics, Bounded Rationality, Learning in Games.
Disciplines: Economics, Political Science, Cognitive Science, Psychology, Organizational Behavior, Computer Science.

Against Game Theory

Introduction

Scholars employ game theory to model interdependent decision-making in bargaining, constitutional law, democratic stability, standard setting, gender roles, social movements, communication, markets, voting, coalition formation, resource allocation, war, and many other domains. For a review, with citations, see (Lucas, McCubbins, & Turner 2013).

Game theory has been widely discredited: many studies have demonstrated game theory's mispredictions. People seem to be highly sensitive to frame or context and biased in their strategies. They follow heuristic decision-making. They are limited in their ability to reason and learn. Many attempts have been made to create a *behavioral* game theory by adding a correction for each of four-dozen odd mispredictions. (See Lucas, McCubbins, & Turner 2013.) But the span of mispredictions is so great and so varied that building in corrections for them produces a model that is computationally intractable for actual human beings.

Review: The Elements of Game Theory

What is a game? A game, theoretically, is defined by identifying the players, the actions available to them, the information they have about the game, when they have such information, what they know about what others know or will know and when those others will know it, the strategies available to them that define rules for what actions they will take in making decisions, the payoffs that come with outcomes, the range of outcomes, and "equilibria." The acronym for these dimensions is PAISPOE: Players, Actions, Information, Strategies, Payoffs, Outcomes, Equilibria. What is an equilibrium? An equilibrium is a path of choices made by players in the game—a path that could happen, given the dispositions of all the players as defined in game theory. An equilibrium concept is a rule a player uses to pursue an equilibrium path. There are a number of proposed equilibrium concepts that players might use. They have names like "Dominant Strategy," "Nash Equilibrium," "Bayesian Strategy," "Correlated Strategy," and "Subgame Perfect Nash Equilibrium (SPNE)." The classic game theory equilibrium concepts are "Dominant Strategy" and "Nash Equilibrium." A pure-strategy Nash equilibrium concept, for example, is a combination of actions for all the players according to which no player can benefit by unilaterally deviating from his or her combination. For a review of types of equilibrium concepts, see (Maschler *et al*. 2013.) To be generalizable, all of the attempts to model interdependent choice, classical or behavioral, must assume that (1) people follow equilibrium strategies, (2) there are specific types of people who choose a generalizable strategy over a class of tasks, or (3) all people in performing a specific task choose a generalizable behavioral strategy. Behavioral game theory is trying to build a layer cake on top of game theory, by adding layers to the original model. Each additional layer consists of a correction to the foundation. For example, Prospect Theory would eliminate certain strategies associated with disfavored bets.

But the number of adjustments needed to build a behavioral game theory is so vast that it cannot yield generalizable models. The literature has proposed many adjustments in terms of bounds, biases, heuristics and context dependencies. This presents two problems for behavioral game theory. First, experimental economists working in the laboratory and knowing that subjects make one or more of these adjustments are no longer in the position of knowing subjects' true payoffs: we know only the experimental economists' view of the experimental earnings, typically thought to be the subjects' earnings. What experimentalists do not consider is "context," that is, factors like unobserved experimenter demand and framing. Second, and most important, under such adjustments, we do not know what the subjects believe. Most games, especially those simple enough to be tested in the lab and simple enough that we can strip out most (if not all) of the effect of framing and context, assume that subjects share *common knowledge*. Classical game theory requires players to have correct and *consistent* beliefs. To have "correct beliefs" is to regard other players as following classical game theory and to predict that they follow classical equilibrium strategies. Indeed, it is also required that players know that other players know that they themselves are following classical strategies and so on, ad infinitum. As Lupia et. al. (2010) note, the condition is even stronger, in that a classical equilibrium concept "requires shared conjectures. … Common Nash refinements ... continue to require that actors share identical conjectures of other players' strategies" (p. 106). This is part of what economists assume when they accept that the players in a game share "common knowledge." As Smith (2000, p. 9) writes, citing two other winners of the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel:

> "The common knowledge assumption underlies all of game theory and much of economic theory. Whatever be the model under discussion . . . the model itself must be assumed common knowledge; otherwise the model is insufficiently specified and the analysis incoherent" (Aumann 1987, p. 473). Without such common knowledge people would fail to reason their way to the solution arrived at cognitively by the theorist. This is echoed by Arrow when he notes that a "monopolist, even . . . where there is just one in the entire economy, has to understand all these [general equilibrium] repercussions . . . has to have a full general equilibrium model of the economy (Arrow 1987, p. 207)."[1]

Nash showed that any two-player zero-sum game has an equilibrium and it was later proved that finding this equilibrium is computationally tractable. (A zero-sum game has fixed payoffs, where higher payoffs to one player result in corresponding lower payments to other players.) This was the appeal of noncooperative game theory: we could find equilibria that in turn predict outcomes of interest. But, even in the case where there are only two players and two actions, we cannot expect humans to solve non-zero-sum games, as they are computationally intractable or infeasible, falling into the class of computational problems identified as PPAD-complete[2] (for history, definition and analysis see Daskalakis et al. 2009; see also Papadimitriou 2005; Chen & Deng 2006).

The core problem for building a behavioral game theory is that, as we add the biases, heuristics, and context dependencies suggested by decades of research, we are implicitly increasing the dimensionality of the computational problem of finding an equilibrium. It may seem that building a range of possible deviations into the model would help us build a better model. Doing so requires the bold assumption that a given subject, when faced, e.g. with a series of non-zero sum games, deviates from classical game theory in a way that is *consistent* from one game to the next, and even from one choice to the next inside the same game. The result is indeed a generalizable theory -- perhaps a false theory, but one that is at least generalizable. But for that theory to give us purchase on modeling human thought or predicting human behavior, the deviations in beliefs and therefore strategies, must—at a minimum—be consistent. They also must be common knowledge: subjects will have no way to compute consistent strategies if the way in which subjects deviate from strategies is private information belonging to only the capricious subject and unknowable to other subjects. That is the bedrock on which the rescue operation for failed game theory must be built. Our evidence, however, like all other evidence of which we are aware that touches on this point, shows that this presumed bedrock does not in fact exist.

New Findings

Research Design

There are two important new features of our experimental design. First, we created a battery of tasks. Our experiments use a battery of up to 17 games, several of which we constructed by modifying the standard form of well-known games such as “Prisoner’s Dilemma,” “Public Goods,” “Stag Hunt,” “Ultimatum,” “Trust,” “Chicken,” “Dictator,” etc. The purpose of our modifications is to minimize or eliminate the framing of the game and to present the games, to the extent possible, as starkly as they are defined in prominent textbooks in game theory. For details, see (Lucas, McCubbins, & Turner 2013). We emphasize that, unlike the typical method of running experiments in psychology and economics, where subjects face the same task repeatedly, our method presents subjects with a battery of tasks. Additionally, unlike psychology experiments, where subjects are typically paid in the form of satisfying a course requirement, and unlike economics experiments, where subjects are typically paid at random for only one of the two to three dozen repetitions of the task, our subjects know that they are paid in cash according to *every* action they take in *every* task.

Second, we added prediction markets, based on (Plott and Roust 2005; Wolfers and Zitzewitz 2004). In these prediction markets, subjects could earn additional money by placing bets on the choices that were made by the players with whom they were matched and in many cases by placing bets on the collective choices of all subjects who had played the game during previous runs of the experiment. We create a market, specifically a betting market, where we invite subjects to bet on other players' choices. We quiz the subjects on the betting procedure so that we could both motivate them to work hard to understand what the bets entailed and also to measure each subject's compliance with respect to the betting tasks. The subjects were paid if their bets regarding the other player's actions were accurate and they were given a chance to also double down on their bets if they felt confident about their predictions. Subjects understood that they can earn this extra money from betting, and how much for each bet. For example, in the Trust Game, there are two players randomly paired—Player 1 and Player 2; both start with $5; Player 1 can select any number of dollars from 0 to 5 to give away; those dollars are taken from Player 1, tripled by the experimenters, and given to Player 2; after receiving them, Player 2 has the opportunity to transfer back any number of the dollars that Player 2 has in total, including Player 2's original $5. At this point, Player 2 may have any number of dollars between 5 and 20. We ask Player 1 to guess how many dollars Player 2 will return. Later, but before Player 2 learns Player 1's choice, we ask Player 2 to guess how many dollars Player 1 selected to give away.

We also ask Player 2 to guess how much Player 1 predicted Player 2 would transfer back to Player 1. After Player 2 learns Player 1's choice, we then ask Player 2 to guess how much Player 1 predicted that Player 2 would return. All players know that a player earns $3 for each correct guess and nothing for a guess that is wrong. The questions we ask vary slightly for each task, but, as an example, here is an exact question we ask Player 1 in the Trust Game: "How much money do you guess the other person transferred back to you? If you guess correctly, you will earn $3. If not, you will neither earn nor lose money." Players know that, with one exception, they can *never* learn whether their predictions (bets) were right or wrong and that subjects *never* have any information about other subjects' guesses (bets). The exception is the rare case in which Player 2 in a sequential game, such as Trust or Ultimatum, must know Player 1's choice in order for Player 2 to understand Player 2's situation and payoffs. For example, Player 2 in Trust must be informed of how many dollars Player 1 chose to give away, if Player 2 is going to know how many dollars Player 2 has and therefore what the possible actions and payoffs are. Even then, the delivery of this information to Player 2 and Player 2's subsequent choice are postponed to the last set of choices a Player makes, so as to have no effect on previous choices. Payments to subjects are made in a lump sum, without accounting or explanation, individually, anonymously and privately, when the experiment is completed.[3] We do this to eliminate any opportunity for subjects to make inferences about other players' choices (either individually or collectively), and subjects know this before they make any of their many choices in the extensive battery. For details on these prediction markets in a range of games, see (Lucas, McCubbins, & Turner 2013).

Results

Across our battery of tasks, our results verify decades of research demonstrating that subjects do not follow game-theoretic predictions. Consider a traditional Ultimatum Game. According to Andreoni and Blanchard (2006, page 307), this game "has come to symbolize the power" of classical game theory and "its utter failure in practice." In this "bargaining game," a "proposer" makes a take-it-or-leave-it offer to a "responder," who then accepts or rejects the offer. The classical equilibrium prediction is that, if players care only about their own monetary payoffs, then the responder will accept any positive offer the "proposer" makes. (More technically, not counting the "endowment" of dollars with which players begin—in our experiment, for example, the proposer begins with $10 and the responder with $0—, if we assume that players care only about the money they earn, then the Nash Equilibrium prediction for the responder is that the responder will accept any positive offer the "proposer" makes; and the Nash Equilibrium prediction for the proposer is that, knowing that the responder will accept any positive offer, the proposer will reason by backward induction to choose the smallest possible positive offer allowed by the game.) Despite the stark framing of our experimental tasks, our results generally replicate what others have found about the poverty of classical Nash Equilibrium predictions (see Camerer 2003). For example, only slightly more than 6% of our subjects chose the classical strategy when they play the Ultimatum Game.

But the question we focus on with our battery of experimental games is, when players "deviate" from the classical strategies, do they do so consistently? The answer is no, and if there is no consistency, then the modifications to classical game theory cannot provide a generalizable model of behavior. For example, the first half of our Trust Game is exactly like a game that we call "Donation": a given subject as Player 1 in Trust is in exactly the same payoff and action situation as he is when he is the Donor in Donation. Accordingly, we can measure whether a given subject is consistent across these two situations. In addition, the second half of our Trust Game is exactly like our Dictator Game: a given subject as Player 2 in our Trust Game faces the same incentives and action possibilities as he faces when he is the Dictator in our Dictator Game.[4] We thus can measure whether a given subject is consistent across these two situations. In addition, all subjects completed a Trust Game where they made choices as Player 1 and also completed a Trust Game where they made choices as Player 2. We therefore can examine the consistency of a given subject's behavior *within a game,* namely Trust. (See Lucas, McCubbins, and Turner 2013 for details.)

Our findings show that in Trust, there is large variance in behavior across subjects in the role of Player 1 and also in the role of Player 2. (We discuss findings in more detail in an appendix.)

They also show that there is large variance in behavior by the same subject in those two roles. We expected to find that subjects deviate from predicted behavior for each and every task, but, for example, our results show much more than that for Trust, Dictator, and Donation: fewer than 15% of our subjects deviate *consistently* from classical equilibrium concepts across the four tasks involved in those games.[5] For these four tasks, by even the most minimal definition of consistency, only 42% of our subjects either consistently follow classical equilibrium concepts *or* consistently deviate from them. That is, a minority of subjects have even the most minimal consistency from task to task. We report on similar results in (Lucas, McCubbins, & Turner 2013).

This suggests that simple amendments to game theory, such as adding social preferences of one sort or another, risk preference, or discovering each subject's individual level of experience or ability to undertake the reasoning necessary to choose an optimal strategy, or the level of randomness in each individual's choices, cannot by themselves explain the inconsistent pattern of choices across the battery of games in our experiment. For example, sometimes the subject may "appear" altruistic, in one choice, and then "appear" greedy in another choice, even though the incentives are identical and are stated identically for the two tasks. Some subjects are altruistic some of the time, other subjects are altruistic most of the time, and some subjects are never altruistic. At best, we need amendments to game theory that differ for each and every game and for each individual. This would make it difficult, or impossible, to provide a general explanation of behavior that is built on game theory.

Game theoretic predictions about behavior depend upon the assumption that beliefs and choices are aligned. Researchers rarely possess knowledge of the actual beliefs of subjects. But our within-subject experiments allow us to test beliefs, and we find that for individual subjects there is routine and ubiquitous inconsistency across choice and beliefs. (See appendix and McCubbins, Turner, & Weller 2012.) We demonstrate that subjects' beliefs are often inconsistent with equilibrium predictions, which has not been widely appreciated. Our findings also show that these deviations are not consistent; they depend on the specific setting and task. These deviations are so pervasive and so various even within a single subject that is seems unwarranted to refer to them as "deviations." On the contrary, consistent "Nash behavior and beliefs" appear to be remarkable deviations from human cognitive patterns and human behavior.

Advance Directive for Behavioral Game Theory: Do Not Resuscitate

Our results show, as has often been shown, that subjects deviate from the predictions made by classical equilibrium analysis in game theory. We emphasize that (1) subjects often deviate from these predictions, but that (2) for the vast majority of subjects, their deviations are themselves not consistent even across similar tasks, and (3) there is large variance in how different subjects choose for any individual task. Moreover, we show (4) that individuals' beliefs about other subject's choices or beliefs do not support classical Nash Equilibrium strategies, and (5) that there is large variance between subjects and among a single subject's beliefs from task to task. Additionally, we show that individuals do not hold common beliefs about game strategy or deviations from equilibrium. Individuals' beliefs seem to be specific to particular settings and not generalizable from one setting to the next. Indeed, it may be misleading to refer to these patterns of action and belief as "deviations" at all. There are no consistent deviations from classical equilibrium concepts, and thus there is no general behavioral fix. There are about four dozen deviations from classical game theoretic predictions identified in the literature. We find that the same individual subject will be deviating from game theoretic predictions in as many ways as there are tasks in our experiment and that across all subjects in our experiment we see a great variability from one subject to the next in the pattern of deviations. It is unsurprising, for example, that some subjects are altruistic in some settings. It is unsurprising that we find that some fraction of the subjects are altruistic for one or another task. It is more surprising that subjects are altruistic in one task and then not altruistic when offered the identical incentives again later on. All these different variations of course interact with each other, giving a complex and unpredictable landscape of complex variation running over individuals and groups. Thus, adding behavioral assumptions to the general model of cognition within game theory cannot make these general models more suitable for predicting behavior.

We have shown further that the protected core of game theory—the unrecognized cognitive model, or Theory of Mind (McCubbins et al., 2012), of non-cooperative game theory—fails repeatedly in hypothesis testing. The assumptions about human cognition that are part of game theory, including the predictions of classical game theory and its refinements, are at odds with what we know about actual human cognition. This is no surprise, because the equilibrium concepts were not constructed based on how actual humans think, reason, or make decisions. We do not yet see a way forward to creating a behavioral game theory that offers meaningfully generalizable predictions. Accordingly, our advance directive would say: Do Not Resuscitate.

The Next Step

What are the alternatives to game theory, both classical and behavioral? What possibilities are there for forming testable hypotheses about interdependent decision-making? [6] Within cognitive science, there are a number of lines of research about how actual human beings make actual decisions, several of them reviewed in (Turner 2001). These lines of research in cognitive science have had virtually no consideration inside economics. Here are a few of them:

1. *Variation across domains and situations*. Entire subfields of cognitive science are dedicated to the ways in which human thought varies across different domains and situations. Given basic considerations of evolutionary development and fitness, there is no reason to assume a priori that the way a human being thinks and acts with respect to food, mating, entertainment, and so on would follow the same patterns or principles of reasoning and decision. All models of game theory, classical and behavioral, assume the opposite, namely, that although people might have different preferences with respect to these different domains, their patterns of reasoning and choice must be uniform across them. Game theory models two situations as identical if they have the same game structure, regardless of the content. It does not matter, for example, that the "Stag Hunt" game might concern growing vegetables or killing soldiers. To a

cognitive scientist, or maybe just to anybody other than a game theorist, that approach looks like a non-starter. It certainly should be discarded as an assumption.

2. *Learning*. Entire subfields of cognitive science are dedicated to the various mental operations involved in learning, and to the study of how human thought and action depend upon the highly flexible and powerful learning for which human beings are equipped. To give one example, cognitive science routinely considers the power of analogy or blending: one remembers a previous specific situation, perhaps in childhood, or from a biography, or even from a science fiction novel, and remembers, too, its outcomes; one then uses those concepts and knowledge to inform one's understanding and decisions in the present specific situation. These traditions in cognitive science have no status inside game theory, even though they provide a basis for hypotheses and tests about decision-making. Indeed, the validity of measuring such powers of analogy and blending is so unquestioned in Western culture that assessments of this ability are used as part of the process of deciding who has what IQ and which applicants to college should be admitted.
3. *Complexity and nonlinearity*. The world is rich, and in the typical situation, actors are engaged in simultaneous games that overlap. In life, any action is usually a move in many different games. Strategies to maximize expected utility over all these games are typically nonlinear. In principle, the output of any subgame of any game can be input to any subgame of any other game. Game theory by

contrast assumes a partitioning of thought and action to tiny scripts of activity that are pretty much separate from all others. This assumption could be discarded.

4. *Adaptive behavior*. In the typical situation, people are adaptive: their first and strongest disposition is often not to play the game but to reinvent it, change it. Their decisions can be driven by attempts to change the game from the outside. Game theory leaves no room for this normal and routine thought and behavior.
5. *Construal*. Cognitive science routinely investigates our rich capacities for differing *construals* of the same given material. “The mountain range runs from Canada to Mexico” and “The mountain range runs from Mexico to Canada” deal with the same stuff but call for quite different emphases and viewpoints. They also both call for conceptualization by using the idea of *motion* (“runs from . . . to”) even though in some sense we think that no motion is involved. Construal is a crucial part of interdependent decision-making, because actors try to reconstrue history and to get other actors to do the same. In the typical situation, actors work at conceptual reinterpretation of the history of play, so as to persuade other actors that the value and status of a past action must be changed, and further, to persuade them that the action led to nodes different from those to which it was once thought to lead. Conceptually, the history of the game is not fixed. Game theory assumes the opposite.
6. *What’s up?* Actors must operate in general without knowing what game they are in, and the question always arises, who has the authority to recognize and establish the game being played? Actors attempt to influence other actors' thoughts about the game being played.

7. *Identity*. Cognitive science routinely considers the work people do to construct an identity for themselves, and to carry it and vary it appropriately from situation to situation. It may often be that the principle payoff in any scripted activity is not the local payoffs but the actor's concept of a personal identity. When we enter different situations, different rooms, different moments, what does the present offer by way of allowing us to construct an identity? What looks like fatal inconsistency from the point of view of game theory might look like fruitful experimentation, learning, and fluidity from the point of view of actual human beings.

Within the social sciences, it has been shown that institutions (laws, constitutions, auction mechanisms, common agency, families, friendships, societal structures, and so on) serve to create not only incentives for choice but also a set of shared mental models about players, actions, payoffs, outcomes, and perhaps most important, information. This would imply that the study of institutions might supply some of the cognitive grounding that game theory is currently lacking. Yet another assumption of game theory is that from knowledge about players, actions, payoffs, outcomes, and information, one can derive strategies and equilibria. It is unknown whether this is true, but what is clear is that institutions have already influenced how subjects derive strategies and equilibria. People, developed within institutions, are thus, when they enter our experiments, far from a tabula rasa. There is little reason to imagine that the narratives that we can give them in an experiment are strong enough to offer much hope of overcoming that training within institutions. There is little reason to think that these narratives in experiments would substitute for the purported “10,000 hours” of learning and practice needed to be successful at a given task. Accordingly, the cognitive study of decision-making must include the study of learning within institutions.

In short, we think that the future for game theory, if there is one, would come from a grounding in cognitive science, or more generally, in the analysis of how the cognitively modern human mind works, what its basic mental operations are, and how they are deployed in situations. We know that strategic games like chess arose very late in human evolution and even in human culture, and that people are very poor at such games in general and must undergo extensive training in order to play them well. Such games of strategy are perhaps the last place one should look for a model of human decision-making generally. Game theory has placed itself into that cul-de-sac, but there is no reason for that sterility and isolation. One could instead begin with how actual human beings think.

Appendix on experimental data

This appendix provides details of subjects' behavior in Trust, Dictator, and Donation. First, examining play within the Trust Game, we find that 56% of subjects as Player 1 sent money to Player 2. On average, they sent $1.43, with a standard deviation of $1.70. On average, as Player 2, they return $1.23, with a standard deviation of $2.29. Our emphasis is not on the well-established deviance from classical predictions in Trust (indeed, the standard deviation from our experiments is somewhat smaller than that usually reported), but rather on the large variance in behavior both across subjects in both tasks and by the same subject across different tasks. Only 1 of the 80 subjects who as Player 2 received $0 from Player 1 returned any money to Player 1. Of the 100 subjects who did receive money as Player 2, 62 of them returned something. The average returned for this subset is $2.22, again with a large variance (the standard deviation is $2.71). Second, for the 62 who as Player 2 returned money to Player 1 after receiving money, only 40 sent money when they were the Dictator; and of those 40, only 29 sent money when they were in the role of Donor. Was a subject's pattern of deviation from classical equilibrium consistent? There are 42 subjects who deviate from classical equilibrium as *both* Player 1 and Player 2 in Trust. Of these 42, 33 also deviated as Donor and, of these 33, 26 deviated as Dictator. We see that fewer than 15% of our subjects *consistently* deviate from classical equilibrium concepts across these four tasks. In sum, by even the most minimal definition of consistency, only 42% of our subjects either consistently follow classical equilibrium concepts (specifically, "Subgame Perfect Nash Equilibrium," SPNE) *or* consistently deviate from them, in these four tasks. That is, a minority of subjects have even the most minimal consistency from task to task. We report on similar results in (Lucas, McCubbins, & Turner 2013).

We turn now to reporting on the often hidden and never-tested parts of "equilibrium concepts" in game theory, that is, the assumptions regarding subjects' beliefs and knowledge. In the Trust Game, for example, the classical SPNE prediction for both players is that they will send $0 for all tasks. Thus, if our subjects hold beliefs that support SPNE, both Player 1 and Player 2 should expect the other person in each task to send $0. Further, they should expect that the other player expects that they will send $0. (And so on ad infinitum: they should expect that the other expects that they expect that the other player will send $0; and so on.) When acting as Trust Player 1, however, 88 of 180 subjects bet, and thus can be thought to believe, that Player 2 will return some money to them. Likewise, when in the role of Player 2 in Trust, the majority of participants (112 of 180) believe that Player 1 will send them *more* than the equilibrium amount of $0. Indeed, only 21% (38 of 180) hold the "correct" SPNE beliefs and, as both Player 1 and Player 2, bet that the other person will send nothing. (More elaborately, since the "other person" is always that other subject with whom the subject has been randomly paired for that specific task, only 21% of players expect both when they are Player 1 and later when they are Player 2 that the other subject with whom they have been randomly paired for that particular game will send nothing.) 58 of 180 participants bet when they are in both roles—that is Player 1 and Player 2—that the other person will send more than $0. In general, participants do not hold beliefs that support a SPNE.

Examining consistency of belief in more depth, we can ask, for example, how many of the 180 subjects in our analysis consistently held beliefs that support a SPNE? As Player 2, participants made guesses about Player 1's prediction of how much Player 2 would return, and only 92 of 180 made guesses that support a SPNE. Of those 92, only 41 also held SPNE-consistent beliefs as Player 2 when guessing how much Player 1 predicted that Player 2 guessed Player 1 would transfer. Of those 41, 33 were also SPNE-consistent as Player 2 when guessing how much Player 1 would transfer. Of those 33, 29 were also SPNE-consistent as Player 1 when guessing how much Player 2 predicted that Player 1 guessed that Player 2 would return. Of those 29, 27 were also SPNE-consistent as Player 1 when guessing both (A) how much Player 2 would return and (B) how much Player 2 predicted that Player 1 would send. In sum, only 15% of our subjects *consistently* adhere to beliefs that support a SPNE as Player 1 and as Player 2 in a single game, Trust.

References

Andreoni, J. and Blanchard, E. (2006) Testing subgame perfection apart from fairness in ultimatum games. *Experimental Economics*. (2006) 9:307-321.

Andreoni, J. and J.H. Miller. (2002) Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, v. 70, no.2, March, 737-753.

Andreoni, J. and J. Rao. (2010). The Power of Asking: How Communication Affects Selfishness, Empathy and Altruism." *Journal of Public Economics.*

Ardalan, K. (1997) Game Theory, Information Economics, Rational Expectation and Efficient Market Hypothesis: Overviews and Exposition of Interconnections. *International J. of Business*, Vol. 1 No. 2.

Aumann, R. (1981). "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, Vol 4 of *Gesellschaft, Recht, Wirtschaft, Wissenschaftsverlag*, edited by V. Bohm, Bibliographisches Institut, Mannheim, pp. 11-42.

Aumann, R. 1987. "Game Theory." *New Palgrave Dictionary of Economics*, vol. 2. New York: Macmillan, 473.

Axelrod, R. and Hamilton, W. D. (1981). The evolution of cooperation. *Science,* v. 211, 1390-1396.

Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer.* Hillsdale, NJ: Erlbaum.

Batson, C. D., Van Lange, P. A. M., Ahmad, N., and Lishner, D. A. (2007). Altruism and helping behavior. In M. A. Hogg and J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 279-295). London: Sage Publications.

Bentham, J. (1789). *The Principles of Morals and Legislation*.

Berg, J. E., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior*, v. *10*, 122-142.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Camerer, C. F. and G. Loewenstein. 2004. Behavioral Economics: Past, Present, Future. In *Advances in Behavioral Economics*, C.F. Camerer, G. Loewenstein and M. Rabin (editors.) 3-51

Cialdini, R., Brown, S., Lewis, B., Luce, C., & Neuberg, S. (1997). Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology, 73,* 481-494.

Chen, X., and Deng, X. (2006). Settling the complexity of 2-player Nash-equilibrium. In *Proceedings of the 47th Annual IEEE symposium on foundations of computer science* (FOCS'06; pp. 261–272). doi: 10.1109/FOCS.2006.69

Cooper, D. J. and J. H. Kagel. In press. "Other-Regarding Preferences: A Selective Survey of Experimental Results." *Handbook of Experimental Economics, Volume 2,* edited by J. H. Kagel and A. E. Roth (editors). Princeton University Press. http://www.econ.ohio-state.edu/kagel/Other%20Regarding_All_2_12_13.pdf

Crawford, V. and J. Sobel. 1982. Strategic Information Transmission. *Econometrica* 50: 1431-1451.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment and uncertainty. *Cognition*, *58*, 1 1Crawford, V. and J. Sobel. 1982. Strategic Information Transmission. *Econometrica* 50: 1431-1451.

Croson, R. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry, v. 45*, 199-216.

Daskalakis, C., Goldberg, P. W., & and Papadimitriou, C. H. (2005). The complexity of computing a Nash equilibrium" (TR05–115). Electronic Colloquium on Computational Complexity. Retrieved from http://eccc.hpi-web.de/eccc-reports/2005/TR05-115

Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. (2009). The complexity of computing a Nash equilibrium. *Communications of the ACM*, *52*(2), 89–97. doi:10.1145/1461928.1461951

Dawes, R. M., Van de Kragt, A. J., and Orbell, J. (1990). Cooperation for the benefit of us – Not me, or my conscience. In J. J. Mansbridge (Ed.), *Beyond self-interest* (pp. 97-110). Chicago: University of Chicago Press.

Dovidio, J. F., Allen, J.,and Schroeder, D. A. (1990). The specificity of empathy-induced helping: Evidence for altruism. *Journal of Personality and Social Psychology,* v. 59*,* 249-260.

Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry, 1 10*, 1 1-11.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., & and Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *Journal of Personality and Social Psychology, 78,* 136–157.

Friedman, J. W. (1971). "A Non-cooperative Equilibrium for Supergames". *Review of Economic Studies* 38 (1): 1–12.

Friedman, J. (1996). *The Rational Choice Controversy: Economic Models of Politics Reconsidered*.

Fudenberg, D. and E. Maskin. (1986) The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54, 3, 533-54.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton, (Eds.), *Subjective probability* (pp. 129–161). New York, Wiley.

Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. Dellarosa Cummins & C. Allen, (Eds.), *The evolution of mind*. New York: Oxford University Press.

Gigerenzer, G., P. Todd, and the ABC Research Group (1999). Simple Heuristics that Make us Smart. Oxford University Press.

Gigerenzer, G. and Selten, R. (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, Mass.: The MIT Press.

Gigerenzer, G. and W. Gaissmaier. (2011). Heuristic Decision Making. *Annu. Rev. Psychol.* 62:451–82.

Graziano, W. G., Habashi, M. M., Sheese, B. E. and Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person X situation perspective, *Journal of Personality and Social Psychology,* v. *93*, 583-599.

Hardesty, L. (2009, November 9). What computer science can teach economics. *MIT News*. Retrieved from http://web.mit.edu/newsoffice/2009/game-theory.html

Hoffman, E., K. McCabe, and V. Smith (2008). Reciprocity in Ultimatum and Dictator Games: An Introduction. In Plott, C. and V. Smith, *Handbook of Experimental Economics Results,* v. 1. Amsterdam: North-Holland: 411-416.

Griffin, D. Gonzalez, R., & Varey, C. (2001). The heuristics and biases approach to judgment under uncertainty. A. Tesser and N. Schwarz (Eds), *Handbook of Social Psychology: Intraindividual Processes* (pp. 207-235), Blackwell: Oxford, UK.

Hobbes, T. (1651). *Leviathan*. Oxford: Oxford University Press.

Jones, S. K., Jones, K. T., & Frisch, D. (1995). Biases of probability assessment: A comparison of frequency and single-case judgments. *Organizational Behavior and Human Decision Processes*, *61*, 1 Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Macmillan.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Macmillan.

Kahneman, D., J. L. Knetsch and R. Thaler. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, v.* 5 193-206.

Kahneman, D. and A. Tversky. (1979). Prospect Theory: An analysis of decision making under risk. *Econometrica,* 47 *2*63-291.

Kuhlman, D. M., and Wimberley, D. L. (1976). Expectations of choice behavior held by cooperators, competitors, andand individualists across four classes of experimental game. *Journal of Personality and Social Psychology, v. 34*, 69-81.

Kramer, R. M., McClintock, C. G., and Messick, D. M. 1986. Social values and cooperative response to a simulated resource conservation crisis. *Journal of Personality, v. 54,* 576-592.

Latané, B., and Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology, 1,* 65–82.

Lucas, G., McCubbins, M. & Turner, M. 2013. "Can We Build Behavioral Game Theory?" Available at SSRN: http://ssrn.com/abstract=2278029 or http://dx.doi.org/10.2139/ssrn.2278029

Luce, D. & Raiffa, H. (1957) *Games and Decisions.* NY: Wiley.

Lupia, A., A. S. Levine, and N. Zharinova. 2010. "Should Political Scientists Use the Self Confirming Equilibrium Concept? Benefits, Costs and an Application to the Jury Theorem." *Political Analysis* 18:103-123.

Maschler, M., Solan, E., & Zamir, S. 2013. *Game Theory*. Cambridge University Press.

May, K. O. (1954). Intransitivity, utility and the aggregation of preference patterns. *Econometrica, v.* 2, 1-13.

McAdams, R. (2009) Beyond the Prisoner's Dilemma: Coordination, Game Theory and the Law. January, 82 S. *Cal. L. Rev*. 209

McCubbins, Mathew D., Mark Turner, and Nicholas Weller. 2012. "The Theory of Minds Within the Theory of Games." In Arabnia, H. R., de la Fuente, D.

Kozerenko, E. G., LaMonica, P. M., Liuzzi, R. A., Olivas, Jose A., Solo, A. M. G., and Waskiewica, T., editors, *Proceedings of the 2012 International Conference on Artificial Intelligence,* Vol. I. CSREA Press, pp. 515-521.

McCubbins, Mathew D., Mark Turner, and Nicholas Weller. 2013. “Testing the Foundations of Quantal Response Equilibrium.” In Greenberg, A. M., Kennedy, W. G., Bos, N. D. (Eds.), *Social Computing, Behavioral-Cultural Modeling and Prediction. Lecture Notes in Computer Science*, Volume 7812. Berlin: Springer. Pages 144-153.

McKelvey, R. and Palfrey, T. 1995. Quantial Response Equilibria for Normal Form Games,” *Games and Economic Behavior*. 10, 6-38

McKenzie, C. R. M., & and Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, v. *54*, 33-61

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Mill, J. S. (1862). The Contest in America. *Harper's New Monthly Magazine, 24(143),* 683-684.

Milinski, M., Semmann, D. and Krambeck, H.-J. (2002). Reputation helps solve the tragedy of the commons. *Nature, v. 415,* 424-426.

North, D. (2005). *Understanding the Process of Economic Change*. Princeton: Princeton University Press.

Ordeshook, P. (1986). *Game Theory and Political Theory: An Introduction*. Cambridge University Press.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., and Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology, v. 56,* 365-392.

Penner, L. A., Fritzsche, B. A., Craiger, J. P., and Freifeld, T. R. (1995). Measuring the prosocial personality. In J. Butcher and C.D. Spielberger, Editors, *Advances in personality assessment* (Vol. 10, pp. 147–164). Erlbaum: Hillsdale, NJ.

Plott, C. and V. Smith. (2008). *Handbook of Experimental Economics Results*. Amsterdam, North-Holland.

Plott, C. and K. Roust. (2005). The Design and Testing of Information Aggregation Mechanisms: a Two-Stage Parimutuel IAM. Social Science Working Paper # 1245. Pasadena, California Institute of Technology. http://www.hss.caltech.edu/SSPapers/sswp1245.pdf

Rapoport, A. and A. M. Chammah. (1965). *Prisoner's Dilemma.* Ann Arbor: University of Michigan Press.

Rasmusen, E. (1992). Folk theorems for the observable implications of repeated games. *Theory and Decision* 32 (2):147-164

Rogers, Brian W., Palfrey, Thomas R. & Camerer, Colin F. 2009. "Heterogeneous quantal response equilibrium and cognitie hierarchies. *Journal of Economic Theory.*Volume 144, issue 4. Pages 1440-1467.

Samuleson, P. (1947). *Foundations of Economic Analysis*. Harvard University Press.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings and the logic of conversation. *Advances in Experimental Social Psychology*, 26, 123-162.

Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativenessand the logic of conversation: The contextual relevance of “irrelevant” information. *Social Cognition*, *9*, 67–84.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84*, 60-70.

Simon, H. (1957). *Models of man: Social and rational*. New York: Wiley.

Smith, K. D., Keating, J. P., and Stotland, E. (1989). Altruism Reconsidered: The Effect of Denying Feedback on a Victim's Status to Empathic Witnesses. *Journal of Personality and Social Psychology,v. 57,* 641-50.

Smith, V. (2000). “Rational Choice: The Contrast Between Economics and Psychology” in *Bargaining and Market Behavior: Essays in Experimental Economics* (pp. 7-24). NY: Cambridge University Press.

Tesser, A. (2000). On the confluence of self-esteem maintenance mechanisms. *Personality and Social Psychology Review, 4*, 2 Thaler, R. and C.Sunstein. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness.* Revised edition. Yale University Press.

Thaler, R. and Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness.* Revised edition. Yale University Press.

Turner, M. 2001. “Choice,” chapter 3 of *Cognitive Dimensions of Social Science.* New York: Oxford University Press.

Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., and Joireman, J. A. (1997). Development of prosocial, individualistic and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, v. 73,* 733-746.

Von Neumann, J. & Morgenstern, O. (1944) *Theory of Games and Economic Behavior.* Wiley, NY.

Weber, M., Kopelman, S., and Messick, D. M. (2004). A conceptual review of social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review,* v. *8*, 281–307.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives.*

Zak, P. (2012). *The Moral Molecule: The Source of Love and Prosperity.The Critical Role of Values in the Economy.* Dutton.

NOTES

---

[1] Smith here quotes a reprint of an original article by Arrow: Arrow, Kenneth J. 1986. "Rationality of Self and Others in an Economic System." *The Journal of Business*, Vol. 59, No. 4, Part 2: The Behavioral Foundations of Economic Theory. pp. S385-S399.

[2] PPAD stands for "Polynomial Parity Argument on Directed Graphs." It is a complexity class regarded as exceptionally difficult. In computational complexity theory, a problem X is called "hard" for a complexity class C if any problem in C can be reduced to X, which implies that no problem in C is harder than X, since a solution to X provides a solution to any problem in C. If X is both hard for C and in C, then X is called "C-complete." A problem that is C-complete is the hardest problem, computationally, in C, or rather, there is no harder problem in C. A PPAD-complete problem is in principle "so hard to calculate that all the computers in the world couldn't find it in the lifetime of the universe" (Hardesty 2009). Accordingly, it is difficult to imagine that human beings playing a game would seek solutions by trying to perform such a computation. "By showing that some common game-theoretical problems are so hard that they'd take the lifetime of the universe to solve, Daskalakis is suggesting that they can't accurately represent what happens in the real world" (Hardesty 2009).

[3] Except for on time show up fees that were paid to all subjects and except for payments for the first quiz relating to general experimental instructions.

[4] Of course, one difference remains, the Trust task is interactive while the Dictator task is not.

[5] That is, in making choices as Player 1 or Player 2 in Trust, in making choices in Dictator or Ultimatum, subjects sometimes deviate and sometimes do not deviate from classical Nash Equilibrium predictions. Those that deviate do not always do so, and few subjects deviate on every choice and few never deviate.

[6] Another way in which behavioral game theory has sought to account for ubiquitous deviations from game-theoretic predictions is to add a random factor to human decision-making. One prominent such approach is “Quantal Response Equilibrium” (see McKelvey & Palfrey, 1995). But it is largely impossible to put Quantal Response Equilibrium to the test as almost any pattern of behavior is consistent with its predictions (see McCubbins, Turner, and Weller 2013), especially when it is expanded to “Heterogenous Quantal Response Equilibrium” (Rogers et al. 2009).