

Can We Build Behavioral Game Theory?

Gale M. Lucas

University of Southern California

Mathew D. McCubbins

University of Southern California

Mark Turner

Case Western Reserve University

Abstract: The way economists and other social scientists model how people make interdependent decisions is through the theory of games. Psychologists and behavioral economists, however, have established many deviations from the predictions of game theory. In response to these findings, a broad movement has arisen to salvage the core of game theory. Extant models of interdependent decision-making try to improve their explanatory domain by adding some corrective terms or limits. We will make the argument that this approach is misguided. For this approach to work, the deviations would have to be *consistent*. Drawing in part on our experimental results, we will argue that deviations from classical models are not consistent for any individual from one task to the next or between individuals for the same task. In turn, the problem of finding an equilibrium strategy is not easier but rather is exponentially more difficult. It does not seem that game theory can be repaired by adding corrective terms (such as consideration of personal characterizes, social norms, heuristic or bias terms, or cognitive limits to choice and learning). In what follows, we describe new methods for interdependent decision-making. Our experimental results show that people do not choose consistently, do not hold consistent beliefs, and do not in general align actions and beliefs. We will show that experimental choices are inconsistent in ways that prevent us from drawing general characterizations of an individual's choices or beliefs or of the general populations' choices and beliefs. A general behavioral game theory seems a distant and, at present, unfulfilled hope.

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.

Can We Build Behavioral Game Theory?

Introduction

The formal modeling of interdependent decision-making according to the principles of expected utility theory is called “game theory,” and it is exceptionally influential in the social sciences. Scholars employ game theory to model interdependent decision-making in bargaining, constitutional law, democratic stability, standard setting, gender roles, social movements (McAdams 2009), communication (Crawford and Sobel 1982), markets (Ardalan 1997), voting, coalition formation, resource allocation, war and many other domains (Ordeshook 1986; Rasmusen 1989).

Game theory has also been widely discredited: many studies have demonstrated game theory’s mispredictions (Camerer 2003, Plott and Smith 2008, especially parts 4 & 7). People seem to be highly sensitive to frame or context and biased in their strategies (Kahneman and Tversky 1979), to follow heuristic decision-making (Gigerenzer and Gaissmaier 2011), or to be limited (or bounded) in their ability to reason and learn (Simon 1957; North 2005; see Plott and Smith 2008). Many attempts have been made to create a *behavioral* game theory by adding a correction for each misprediction (Camerer 2003, Plott and Smith 2008). We argue that modeling limitations in learning or computational ability will not ultimately be fruitful and that the span of mispredictions is so great and so varied that the problem of building corrections for them into models is computationally intractable.

Attempts to build theories that match observations

There have been many attempts compensate for the failure of game theory to match observed behavior. For example, Allais (1953) surveyed a wide range of game-theoretic mispredictions. In response, Kahneman and Tversky (1979), among others, attempted to explain such mismatches as arising from consistent biases, such as the “certainty effect,” captured in “Prospect Theory.” Other findings that needed to be explained away included that people do not consistently order their preferences transitively (May 1954) and that cognition is bounded in ways not anticipated in game theory (Simon 1957). Researchers often study such mismatches between expectation and observation in the hope of resolving them by hypothesizing psychological causes—“biases” (such as “projection bias,” “confirmatory bias,” and “status quo bias”; see Camerer and Loewenstein 2004; Thaler and Sunstein 2008). Of course, it could be that the theoretical expectations were misguided in the first place. Gigerenzer, for example, argues that heuristics can save time and effort and can thus lead often to judgments that are more broadly rational (Gigerenzer 1999; Gigerenzer & Selten 2001).

Recent laboratory studies report that people often fail to maximize their own earnings in game-theoretic settings (Smith 2000; Camerer 2003, Berg et al. 1995), where subjects face decisions in which one subject's gain entails loss to another subject or subjects, including two-player games—such as Ultimatum, Trust, Dictator and Prisoner's Dilemma—and multiplayer or social games, such as step-level Public Goods games. It has been claimed that they are biased to be "other-regarding" (Cooper and Kagel in press; Batson, Van Lange, Ahmad, & Lishner 2007; Dovidio, Allen, & Schroeder 1990; Batson 1998; Weber, Kopelman, & Messick 2004). Recent research has suggested that there is a wide variety of personality traits (Graziano, Habashi, Sheese, and Tobin 2007; Penner, Fritzsche, Craiger, and Freifeld 1995; Van Lange, Otten, De Bruin, and Joireman 1997), social contingencies (Latané and Darley 1970) and interpersonal experiences (Batson 1991; Smith, Keating, and Stotland 1989) that inspire people to prioritize other's outcomes over their own (for reviews, see Batson et al. 2007; Penner, Dovidio, Piliavin, and Schroeder 2005). Evidence exists outside the lab for the deviation of observed behavior from game-theoretic predictions (e.g., see Porter 1983b). Additionally, laboratory scientists from all fields have demonstrated that context matters: games that appear to offer identical choices result in different behavior when they are framed differently or when they are conducted in different contexts (e.g., Kahneman and Tversky 1979).

Models meant to explain such deviations of observation from game-theoretic prediction have been proposed (for a survey see the essays in Friedman 1996). In what are perhaps the best-known models, cooperation emerges from indefinitely repeated play (Aumann 1981; Axelrod and Hamilton 1981; Friedman 1971; Fudenberg and Maskin 1986; Rasmusen 1992; Rappaport and Chammah 1965). Other scholars argue that utility functions should be more broadly defined and should take into account other people's outcome or even the outcome for society as a whole (e.g. Andreoni and Miller 2002; Andreoni and Rao 2010; Milinski, Semmann, and Krambeck 2002; Nowak and Sigmund 1998; Dawes, Van de Kragt, & Orbell 1990; Kramer, McClintock, and Messick 1986; Smith et al. 1989).

Can We Create a Behavioral Game Theory?

Defining a game requires specifying *PAISPOE* (see Rasmusen 1989): players; actions; information; strategies; payoffs; outcomes; and equilibria. Overall, in equilibrium, an individual can receive no incremental benefit from changing actions, assuming other players remain constant in their strategies. An equilibrium concept is a rule that defines an equilibrium based on the possible strategy combinations for each player and their payoff functions. Common equilibrium concepts used in game theory include dominant strategy, Nash (NE), Bayesian, correlated strategy, and subgame perfect Nash (SPNE). To be generalizable, all of the attempts to model interdependent choice, behavioral or theoretic, must assume that (1) people follow equilibrium strategies or, (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 non-NE strategy (i.e., a 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 with disfavored bets (Kahneman 2011).

But the number of adjustments needed to build a behavioral game theory is too large to 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, under such adjustments, experimentalists are no longer in the position of knowing subjects' true payoffs: we know only the experimentalists' view of the experimental earnings, typically thought to be the subjects' earnings. What experimentalists don't consider are factors like unobserved experimenter demand, framing, and the like (henceforth referred to as context). Second, and most important, under such adjustments, we don't 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 context, are going to assume that subjects share *common knowledge*. That is, NE requires players to have correct and *consistent* beliefs. To have "correct beliefs" is to regard other players as "Nash players" and to predict that they follow Nash equilibrium strategies. Indeed, it is also required that players know that other players know that they themselves are following NE strategies and so on, ad infinitum. As Lupia et. al. (2010) note, the condition is even stronger, in that NE "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, 9) writes,

"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)

Nash showed that any two-player zero-sum game has an equilibrium and it was later proved that finding this equilibrium is computationally tractable. This was the appeal of noncooperative game theory: we could find equilibria that in turn predict outcomes of interest. But even with only two players and two actions, non-zero sum games cannot be solved; they are considered to be computationally intractable, falling into the class of computational problems identified as PPAD-complete (for history, definition and analysis see Daskalakis et al. 2009; see also Papadimitriou

2005; Chen & Deng 2006). Because NE of non-zero sum games cannot be solved, NE seems unlikely to represent real-world decision-making by inherently cognitively-limited human beings

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. But, if we make the bold assumption that all subjects deviate in *exactly* the same way and do so *consistently* from one choice to the next, the result would be a generalizable theory— perhaps false, but at least generalizable. For this to be true, however, the deviations in beliefs and therefore strategies must—at a minimum—be consistent and be common knowledge. That is the bedrock on which attempts to amend game theory must be built. Our evidence, however, undermines the expectation of consistency in beliefs and choices.

New Findings

Research Design

Our experiment uses a battery of 17 games, several of which are our modifications of the standard form of these games and other related tasks: Prisoner's Dilemma; Five-Player Beauty Contest; Two-Player, Four-Action Beauty Contest; Donation; Matching Pennies (twice); Majority Step-Level Public Goods; Unanimous Step-Level Public Goods; traditional Public Goods; Stag Hunt; Simultaneous Ultimatum (two rounds); Sequential Ultimatum; Ultimatum Regret (for both Players 1 and 2); Trust (closely follows the protocols used by Berg et al. 1995); Simultaneous Discoordination; Chicken; and Dictator (For definitions, see www.mccubbins.us/experiments/battery).

Subjects were recruited using flyers and emails distributed across the University of California, San Diego, and the experiments were conducted throughout 2011. Subjects were not compelled to participate and were given \$5 in cash when they arrived. A total of 200 subjects, drawn at random from our subject pool, participated in this experiment (some tasks have fewer than 200 valid responses). The experiment lasted approximately two hours, and subjects received on average about \$41 in cash.

Subjects arrived in groups of 20 and were divided evenly into two rooms. They were seated behind dividers so they could not see or communicate with each other, or be seen by the experimenters. Tasks were completed using pen and paper, with each task on a separate sheet of paper. Subjects were told that they could make money for every choice they made. Each subject was given a subject number, and subjects were told that their choices were private and anonymous; that we had set up separate accounts for them identified only by their subject number; that, for each task, they were randomly matched with a person (or persons) in the other room, so it is unlikely they would be

matched with the same person from task to task; that they would never know who they were matched with; and that they would never know the earnings for any pairing. Subjects, called by subject number, were paid in private at the end of the experiment. Indeed, until the end of the experiment, subjects received no feedback on their game-theoretic choices.

In general, as game theory assumes no “framing,” our narratives merely describe *PAIP* from *PAISPOE* for each task, leaving it to subjects to discover strategies and to calculate expected outcomes and equilibria. Further, we referred to subjects as “participants,” to the player with whom they were paired as “the person” (or “persons”) “in the other room,” and to all activities as “tasks.” Subjects received written instructions, with clarifying graphics. To create the conditions for common knowledge presumed by NE, subjects needed to know that all subjects had the same information. Thus, a research assistant in each room read the instructions aloud and informed subjects that the same instructions were being read to the subjects in the other room. To measure compliance and to encourage subjects to pay attention, subjects were given paid quizzes throughout the experiment. We quizzed them on *PAIP*. A research assistant collected and scored each quiz, on the spot, circling any wrong answers (which were very few, except for the classic five-player Beauty Contest) so that the subject could see the correct answer to each quiz question. The second play of the *sequential* two-player games all took place at the end of the experiment, so as to limit learning or other potential interactive effects (e.g., group reputation) of the experimental design. We believe this allows us to examine the outcome of each game as separate from the preceding experimental history. The subjects’ forms were collected and shuttled to a third room for real time input into a computer, and for subsequent delivery of the choice made by Trust Player 1 in one room to the Trust Player 2 in the other room with whom Trust Player 1 was randomly paired for that task. We have tested and rejected simple linear models of learning in these tasks.

The most important new feature of our experiment design was that we added prediction markets (Plott and Roust 2005; Wolfers and Zitzewitz 2004), wherein subjects could earn additional money by placing bets on the choices that were made by the players with whom they were matched. In essence, the subjects pick the terms of the bet and we sell the bet. For the Trust game, for example, after Player 1 (P1) makes his choice about how much to pass, we ask P1 to guess how much Player 2 (P2) will return. Later, but before P2 learns P1’s choice, we ask P2 to guess how much money P1 passed. We also ask P2 to guess how much P1 predicted P2 would transfer. After P2 learns P1’s choice, we then ask P2 to guess how much P1 predicted she would return. All players know that all players earn \$3 for each correct guess and they earn nothing for a guess that is wrong. The questions we ask vary slightly for each task, but, as an example, here is the exact question we ask P2 in Trust: “How much money do you guess the other person chose to transfer? If you guess correctly, you will earn \$3. If not, you will neither earn nor lose money.” Players are told that they can never learn whether their predictions were right or wrong and subjects never have any

information about other subjects' guesses. Previous research has found that people often hold beliefs about other players that are inconsistent with Rational Choice Models and NE (Kuhlman and Wimberely 1976; Croson 2007).

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), "The ultimatum game (Guth et al., 1982) has come to symbolize the power of subgame perfection in game theory and its utter failure in practice. In this bargaining game, a 'proposer' makes a take-it-or-leave-it offer to the 'responder,' who subsequently accepts or rejects the offer. The subgame perfect equilibrium prediction if players care only about their own monetary payoffs is that the proposer offers nothing (or almost nothing) to the responder, who accepts any positive offer" (p. 307). Our results generally replicate what others have found with respect to NE predictions (Plott and Smith 2008, part 4.1). For example, only slightly more than 6% of our subjects chose the SPNE strategy in the Ultimatum Game.

For brevity, we now leave the topic of the accuracy of NE predictions and move on to the question of consistency of players' actions. In our Trust protocol, two players (P1 and P2) each begin with a \$5 endowment and P1 chooses how many dollars, if any, to pass in private to an anonymous P2, who receives three times the amount transferred. This first half of the Trust Game, when played alone, defines our Donation Game (in Donation P1 is called the Donor and P2 has no choice to make). In the Trust Game, P2 then must choose how much, if any, of the total amount (the sum of P2's \$5 initial endowment plus three times the amount P1 transferred) to return to the first player (borrowed from Kahneman et. al 1991; Zak 2012). The second half of the Trust Game is recognized to be a Dictator Game, with P2 being the Dictator and P1 being the Receiver. Dictator is like Donation: the Dictator is given an endowment and chooses whether or not to transfer some of this endowment to the Receiver, who has no choice to make. In our version of Dictator, we establish the endowments to be exactly the same, for the Dictator, as the endowments that subject faced as P2 in Trust, although the pairing for the Dictator is again another randomly drawn subject. Thus, each participant is given four tasks in our Trust Game complex (Donor, Trust P1, Trust P2, Dictator; they are also the Receiver in Donor and Dictator, but have no task to perform). For each game, SPNE, which is also a dominant strategy equilibrium, for all players equals \$0. 48 of 180 (27%) subjects take actions that are consistent with SPNE across all 4 of these tasks. In our battery, subjects understand that they are randomly and anonymously paired anew with someone for each task.

Because all subjects completed a Trust game where they made choices as P1 and one where they made choices as P2, we can also examine the consistency of their behavior *within a game*. We will show that subjects' behavior in

Trust, Dictator, and Donation are inconsistent. We do so in three steps. First, examining play within the Trust game, we find that 56% of subjects as P1 sent money. On average, they sent \$1.43 (s.d. \$1.70). On average, as P2, they return \$1.23 (s.d. \$2.29). Our emphasis is not on the well-established deviance from SPNE in Trust, 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 received \$0 as P2 returned any money. Of the 100 subjects who did receive money as P2, 62 of them returned something. The average returned for this subset is \$2.22, again with a large variance (s.d. \$2.71). Second, consider the 62 who returned money after receiving money: Did they consistently take money from their account and give it to an anonymous “participant in the other room,” despite the dictates of SPNE? The answer is no: of those 62, only 40 sent money as Dictator; and of those 40, only 29 sent money as Donor. Was a subject’s pattern of deviation from SPNE consistent? There are 42 subjects who deviate from SPNE as both P1 and P2 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 SPNE across these four tasks. In sum, by even the most minimal definition of consistency, only 42% of our subjects are consistent across these four tasks, that is, either consistently following SPNE *or* consistently deviate from SPNE.

Others have demonstrated inconsistency by showing that, for example, choices in the dictator game do not predict choices in an ultimatum game (Hoffman et al. 2008). As described above, our tasks were designed to be *directly* comparable in this way. Therefore, we can model a subject’s choice as P2 in the Trust game using the subject’s choices in the Trust Game Complex. To predict the amount P2 returns to P1 in Trust, we use a linear regression and include the following predictors: the amount P2 received from P1, the amount the subject passed as P1 in Trust, the amount the subject passed in the Dictator game, and the amount the subject passed in the Donation game. The results of this regression indicate that (1) the amount a subject received as Trust P2, (2) the amount a subject passed as P1, and (3) the amount a subject passed as Dictator are all significant predictors of a subject’s choice as P2 in Trust (the amount passed as Donor is not uniquely significant). We can therefore gain some leverage in predicting a subject’s behavior by observing the same subject’s choices in nearly identical tasks within the same two-hour sitting. However, even after we utilize information from the other choices, we are left with more than 40% of the variance still unexplained.

Do You Believe in Magic?

Smith (2000: 9) argues that “optimal market outcomes” must be achieved by “conscious cognition; it can’t occur by magic... .” We agree. We turn now to examine the often hidden and never tested parts of game theory, the assumptions regarding subjects’ beliefs and knowledge. Consider again the Trust Game: , since the SPNE choice is to send \$0 for all tasks, subjects who hold beliefs that support an SPNE should, as both P1 and P2, expect the other

person to send \$0. When acting as P1, however, 88 of 180 participants bet, and thus can be thought to believe, that P2 will return some money. Likewise, as P2, the majority (112 of 180) of participants believe that P1 will send them more than \$0. Indeed, only 21% (38 of 180) hold the “correct” SPNE beliefs and bet that the other person will send nothing as both P1 *and* P2. 58 of 180 participants believe—as both P1 and P2—that the other person will send more than \$0. In general, participants do not hold beliefs that support a SPNE.

We now move on to the question of consistency of players’ beliefs. How many of the 180 subjects in our analysis consistently held beliefs that support a SPNE? For sequential games, SPNE are derived by backward induction. Recall that NE requires players to have an infinite level of correct recursive beliefs. Therefore, irrespective of the level of recursion that we are asking the subjects to bet on, if people believe that others would maximize their own earnings, then they should consistently guess \$0 and buy the bet. We do not find this, however. As P2, participants made guesses about P1’s prediction of how much P2 would return, and only 92 of 180 made guesses that support a SPNE. Of those 92, only 41 were also SPNE as P2 when guessing how much P1 predicted that P2 guessed P1 would transfer. Of those 41, 33 were also SPNE as P2 when guessing how much P1 would transfer. Of those 33, 29 were also SPNE as P1 when guessing how much P2 predicted that P1 guessed that P2 would return. Of those 29, 27 were also SPNE as P1 when guessing both (A) how much P2 would return and (B) how much P2 predicted that P1 would send. In sum, only 15% of our subjects *consistently* adhere to beliefs that support a SPNE as P1 and P2 in Trust games.

If we look at both actions and beliefs in the Trust Game, we find that only five percent of subjects unvaryingly made choices consistent with SPNE and held beliefs (and subsequently bets) that support a SPNE. Our data also show routine inconsistency of choices with beliefs.

It seems that, across various analyses and comparisons, there are consistent inconsistencies across choices, across beliefs, and between choices and beliefs. We find little hope for a general model for behavioral game theory. Can we build behavioral game theory? Perhaps, but not by adding dimensionality.

Discussion

Our results show, as has often been shown, that subjects deviate from NE predictions. More important, we emphasize that (1) for the vast majority of subjects, deviations from NE are themselves not consistent even across similar tasks, and (2) there is large variance in choice between different subjects in any individual task. Moreover, we show that individuals do not hold shared beliefs about game strategy. 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. We do not see how one would build a behavioral game theory that offers meaningful and generalizable predictions.

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 NE 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.

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, 1981, pp. 11-42.
- 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.
- 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

- 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.
- 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. (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.
- 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., 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.
- 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.
- 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*, 47 263-291.
- Kuhlman, D. M., and Wimberley, D. L. (1976). Expectations of choice behavior held by cooperators, competitors, and 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.
- 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.
- 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
- Samuleson, P. (1947). *Foundations of Economic Analysis*. Harvard University Press.
- 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.
- Thaler, R. and C. Sunstein. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Revised edition. Yale 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., & Zietz, 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.

Biographies

Gale M. Lucas is currently a postdoctoral scholar at the University of Southern California.

Mathew D. McCubbins is the Provost Professor of Business, Law and Political Economy at University of Southern California.

Mark Turner is Institute Professor and Professor of Cognitive Science at Case Western Reserve University.