The Mythology of Game Theory

Mathew D. McCubbins[1], Mark Turner[2], and Nick Weller[3]

[1] University of Southern California, Marshall School of Business, Gould School of Law and Department of Political Science
[2] Case Western Reserve University, Department of Cognitive Science
[3] University of Southern California, Department of Political Science

Abstract. Non-cooperative game theory is at its heart a theory of cognition, specifically a theory of how decisions are made. Game theory's leverage is that we can design different payoffs, settings, player arrays, action possibilities, and information structures, and that these differences lead to different strategies, outcomes, and equilibria. It is well-known that, in experimental settings, people do not adopt the predicted strategies, outcomes, and equilibria. The standard response to this mismatch of prediction and observation is to add various psychological axioms to the game-theoretic framework. Regardless of the differing specific proposals and results, game theory uniformly makes certain cognitive assumptions that seem rarely to be acknowledged, much less interrogated. Indeed, it is not widely understood that game theory is essentially a cognitive theory. Here, we interrogate those cognitive assumptions. We do more than reject specific predictions from specific games. More broadly, we reject the underlying cognitive model implicitly assumed by game theory.

Keywords: game theory, human behavior, Nash equilibrium, economics, Trust game, prediction markets

1 The Mythology of Non-cooperative Games

Game theory is essentially cognitive: it assumes that people know the actions available to them, that they have preferences over all possible outcomes, and that they have beliefs about how other players will choose. It assumes that this structure of knowledge, preferences, and beliefs is causal for their choices. Game theory dictates the direction and the patterns of relationships across preferences, beliefs, and choices. Human actions, according to game theory, leap forth fully-formed from our preferences, beliefs, and knowledge of the structure of the game, according to the patterns of execution assumed by game theory.

Researchers have already found that behavior in experimental settings do not accord with these predictions derived from game theory [1]. To explain these discrepancies between theory and behavior, scholars have pointed to (1) cognitive biases and dysfunctions in the decision-making of players [2, 3]; (2) mismatches between a game's payoffs and an individual's utility [4, 5]; and (3) the effects of uncertainty, bounded search ability, or limitations in thinking about others' likely behavior [6, 7, 8, 9]. To be sure, people regularly make choices that do not comport with Nash equilibrium strategies. But this does not necessarily indicate dysfunction: humans solve many tasks that are cognitively quite difficult, and they show great flexibility in their approaches to these tasks [10, 11, 12]. To predict human behavior we must begin with how we actually reason [13], which must be discovered, not assumed. As game-theoretic models are increasingly used across many domains to address important problems [14, 15], it is important to correct these assumptions. These assumptions are pointed out in [16], quoted below. (See also [17]).

> "[T]he Nash equilibrium (NE) concept . . . entails the assumption that all players think in a very similar manner when assessing one another's strategies. In a NE, all players in a game base their strategies not only on knowledge of the game's structure but also on *identical conjectures about what all other players will do.* The NE criterion pertains to whether each player is choosing a strategy that is a best response to a shared conjecture about the strategies of all players. A set of strategies satisfies the criterion when all player strategies are best responses to the shared conjecture. In many widely used refinements of the NE concept, such as subgame perfection and perfect Bayesian, the inferential criteria also require players to have shared, or at least very similar, conjectures." [16: 103-104].

The mythology of game theory is that these identical conjectures spring forward automatically and reliably in all situations, across all settings, consistently for all actors. This is the core that is protected in game theory.

Yet, there is prior work on subjects' beliefs in experimental settings suggesting that subjects in fact do not hold the conjectures and beliefs assumed for them in game theoretic mythology [18, 19, 20]. In what follows, using a within-subjects design across a large battery of common experimental games, we investigate choices, beliefs, and their relationship, if any. It is to a description of our experiments that we turn.

2 Experimental Design

When subjects show up, they are divided into two rooms of 10 subjects each, and seated behind dividers so they cannot see or communicate with each other. Each subject is randomly paired with a subject in the other room. They complete the tasks using pen and paper. Subjects were recruited using flyers and email messages distributed across a large public California university and were not compelled to participate in the experiment, although they were given $5 in cash when they showed up. A total of 180 subjects participated in this experiment. The experiment lasted approximately two hours, and subjects received on average about $41 in cash.

We report on a portion of our battery of experimental tasks derived from the canonical Trust game involving two players [21]. For the Trust game, each player begins with a $5 endowment. The first player chooses how many dollars, if any, to pass to an anonymous second player. The first player keeps any money he does not pass. The money that is passed is tripled in value and the second player receives the tripled amount. The second player then has the initial $5 plus three times the amount the first player passed, and decides how much, if any, of that total amount to return the first player. This is common knowledge for the subjects and they know that their choices are private and anonymous with respect to the other player and to the experimenters. The subgame perfect Nash equilibrium (SPNE) is that Player 1 will send \$0 and Player 2 will return \$0. This is also a dominant strategy equilibrium.

Equilibrium strategies derive from assumed beliefs: the assumption is that all players maximize economic payoffs and believe that all other players do the same. In the trust game, for example, a Player 1 with these beliefs concludes that Player 2 will return nothing and so, as a maximizer, sends nothing. The beliefs that players hold about other players lead to the belief, at every level of recursion, that all players will send \$0, will guess that others will send \$0, will guess that others will predict that everyone will send \$0, and so on ad infinitum.

But what if a subject who does have these assumed Nash beliefs finds himself off the equilibrium path? In the Trust game, only Player 2 could be required to make a choice after finding himself presented with an off-the-equilibrium-path choice. If Player 2 receives any money, the SPNE strategy is still to send \$0 back.

A novel feature of our experimental battery borrows from the idea of a prediction market [22]. We add elements to the basic Trust game in order to tap into subjects' beliefs and conjectures. We ask subjects to to "guess" other subjects' choices, or to guess other subjects' "predictions." We do not ask subjects to report their expectations or beliefs, because asking for a report might have normative implications. In general, we try to provide little or no framing of the experimental tasks offered to our subjects. After Player 1 makes his choice about how much to pass, we ask him to guess how much Player 2 will return and to guess how much Player 2 predicted that Player 1 would transfer. Before Player 2 learns Player 1's choice, we ask Player 2 to guess how much money Player 1 passed. We also ask Player 2 to guess how much Player 1 predicted she would transfer. After Player 2 learns Player 1's choice, we ask Player 2 to guess how much Player 1 predicted she would return. All players know that all players earn \$3 for each correct guess and earn nothing for a wrong guess.

The questions we ask vary for each task, but as an example, here is an exact question we ask Player 2: "How much money do you guess the other person transferred to you? If you guess correctly, you will earn \$3. If not, you will neither earn nor lose money." Players do not learn whether their predictions were right or wrong and subjects never have any information about other subjects' guesses.

Players in the Trust game know they are randomly paired with another subject in a different room. Later in the experiment, all subjects also make choices as Player 2, randomly assigned to a player in the other room who was Player 1. All subjects first make choices as Player 1 and then, roughly 90 minutes later, make choices as Player 2. Subjects never learn the consequences of their actions as Player 1, but of course, when the subject is Player 2, the subject can infer the consequences of her choice.

Subjects also make decisions in a variety of other games, including a Dictator game and what we call the Donation game. In both these games, each subject is randomly rematched with another subject in another room. In the Dictator game, there are two players: the Dictator and the Receiver. It is arranged that the Dictator has the same endowment he or she has in the role of Player 2 in Trust, and that the Receiver has the same endowment he or she has in the role of Player 1 in Trust. These endowments are common knowledge. Accordingly, the Dictator game is identical to the second half of the Trust game. In effect, each subject plays the second half of the Trust game twice, but only once with the reciprocity frame. The SPNE is for the Dictator to send $0 to the Receiver.

The Donation game has two players, a Donor and a Receiver. It is identical to the Dictator game, except that both players start with a $5 endowment, and any money sent by the Donor is quadrupled before it is given to the Receiver. This places the Donor in the same strategic setting as Player 1 in Trust, because the obvious dominant strategy for Player 2 in Trust is to return $0. The SPNE is for the Donor to send $0.

At the end of the battery, we present the subjects with those few tasks that would allow them to learn something about the choices made by subjects in the other room. Subjects are asked to make their choice as Player 2 in the Trust game as one of these final tasks. But in no case do subjects get feedback on their choices.

In what follows, we show that subjects do not behave according to NE, that even in deviating they do not deviate consistently and that they do not hold beliefs that are consistent across similar tasks. We do not see identical conjectures across subjects or anything remotely in that vein.

3 Inconsistent Behavior Within and Across Games

The standard approaches to explaining departures from NE strategies (other-regarding preferences, cognitive constraints, or decision-making biases) implicitly assume that players deviate from game-theoretical expectations in consistent ways. For example, if players prefer to reduce inequality, that preference should be stable across all manner of economic games. If players cannot detect and reject dominated strategies or cannot perform iterated deletion of dominated strategies and then reassess, then they cannot reach a dominant-strategy equilibrium. If they cannot perform backward induction, then they also cannot reach SPNE. Such handicaps should operate in all game environments of equal difficulty. We will show that subjects' behavior in Trust, Dictator, and Donation are strongly inconsistent. We do so in three steps.

First, examining play within the Trust game, we find that 56% of subjects as Player 1 send money. On average, they send $1.43 (s.d. $1.70). On average, in the role of Player 2, they return $1.23 (s.d. $2.29). Such results are well-documented in the literature [3]. Our emphasis is not on the well-known deviance from SPNE, but rather on the large variance in behavior both across subjects in a specific task and by the identical subject across different tasks. This variance casts doubt on the prospects for a single, simple explanation for people's behavior.

Second, of the 100 subjects who as Player 2 receive money, only 62 of them return any money. The average returned is $2.22, again with a large variance (s.d. $2.71). Let's follow the 62 who return money after receiving money. We might expect them to be consistent in sending money when SPNE dictates that they not. But of those 62, only 40 send money when they are in the role of Dictator, and of those 40, only 29 send money when they are in the role of Donor.

Third, there are 60 subjects who behave consistently with SPNE in both roles in the Trust game. Since Dictator and Donation lack the reciprocity frame of Trust, we might expect them all to play SPNE in Dictator and Donation. Indeed, of these 60, 57 send $0 in Dictator, and of these 57, 48 send $0 in the Donation game. In short, 20% of the subjects who play SPNE in both roles in Trust deviate from SPNE within Dictator and Donation. Overall fewer than 27% of our 180 subjects consistently play SPNE across these four tasks.

Alternatively, we might expect that a subject's pattern of deviation from SPNE to be consistent. There are 42 subjects who deviate from SPNE in both roles in Trust. Of these 42, only 33 send money in Donation, and of these 33, only 26 send money in Dictator. We see that fewer than 15% of our subjects consistently deviate from SPNE across these four tasks.

Only 41% of our subjects either consistently follow SPNE in these four tasks or consistently deviate from SPNE in these four tasks. Our subjects do not rigidly follow or deviate from SPNE strategies.

4 Are Beliefs and Behavior Consistent?

Cognitive science gives us considerable reason to doubt that players will behave or hold beliefs identically across different environments, because changes in environments lead to changes in mental activation, which affect behavior and beliefs. As Sherrington famously wrote, the state of the brain is always shifting, "a dissolving pattern, always a meaningful pattern, though never an abiding one" [23]. If the particular tasks induce different states of mental activation, then belief and behavior may well vary accordingly. We have just shown that most subjects are not consistent in their choices. We now show that they do not hold consistent beliefs.

Although to our knowledge it has not previously been done, it is easy to take the strategy that is predicted by NE and see whether players believe that other players will follow the NE strategy. In the Trust game, subjects make guesses as Player 1 about the behavior of Player 2 and, likewise, as Player 2 about the behavior of Player 1. As Player 1, subjects guessed what Player 2 would return, and as Player 2, they guessed what Player 1 would send. Only 38 of 180 guessed both times that the other player would send $0. In other words, only 21% of our subjects have NE beliefs inside just the Trust game. We find that there are 54 subjects who possess NE beliefs as Player 1 but not as Player 2 and 30 subjects who possess NE beliefs as Player 2 but not as Player 1. The overwhelming majority of subjects deviate from NE beliefs during even this single experimental game.

We can compare subjects' beliefs about others in one part of the Trust game with their choices in that same part of the Trust game. For example, we can examine the difference between what a subject chooses to do as Player 1 in the Trust game and what as Player 2 they believe Player 1 will do. The modal category is subjects who believe that other subjects will play like them: 109 of the 180 subjects guess that the choice of the Player 1 with whom they are randomly matched will be the same as their own choice when they were Player 1. Perhaps most surprising, there is a large variance, with 71 subjects (39%) making guesses that differ from their own choices. Their conjectures about what others will do and believe in a situation do not match what they do and believe in the same situation. It is difficult to see how a notion of shared, identical conjectures across players can withstand such a result.

We now return to the 60 subjects who chose $0 as both Player 1 and Player 2 in Trust. We will call them "fully Nash actors" in Trust. We examine here whether their beliefs are also "fully Nash" in the Trust game and whether their actions are "fully Nash" in the related Donation and Dictator games. Of these 60 subjects, 56 of them guess as Player 1 that Player 2 will return nothing, which is consistent with SPNE. We also ask Player 1 to guess how much he or she believes the other player (Player 2) will guess that Player 1 is sending to them. In this task, only 40 of the 60 (66%) subjects guessed that the other player would predict that $0 would be sent. In addition, we ask Player 1 to guess how much Player 2 will predict that Player 1 guesses Player 2 will return and in this task 49 of the 60 subjects (81%) have beliefs consistent with SPNE. Even among these 60 subjects, the percent that have SPNE-consistent beliefs varies across questions, further demonstrating that subjects do not have rigid beliefs.

We now turn to a different part of the mythology of game theory having to do with beliefs and behavior. The mythology of game theory assumes that people enter every setting with preferences over outcomes and with beliefs and conjectures about how other players will act and what they will believe. In this mythology, it cannot make a difference whether one asks them to choose an action in the game before making a prediction or the reverse. To interrogate this mythology, we report [24] on results from a unanimous Public Goods game. In this game, each subject is paired with 9 other subjects. Each player has a $5 endowment, and is asked whether they wish to contribute this $5 to a pot or withhold it. The contributors lose that $5 and the non-contributors keep it. If they all contribute, each receives $15. If fewer than ten contribute, then each receives nothing.

In some cases, we ask the subjects first to choose whether to contribute and second to guess the number of other subjects who will contribute. In other cases, we present the tasks in the reverse order. Our experiments show that subjects who choose first guess on average that 3.3 other players will contribute, while subjects who guess first guess on average that 4.6 other players will contribute ($p=0.03$ in a Kolmogorov-Smirnov equality of distributions test), with 80 subjects in each group. Further, in an equality of proportions tests, 25% of subject choose to contribute when making their choice before their prediction, whereas 43% choose to put their money in the pot when prompted about their beliefs before they made their choice ($p<.03$). This result suggests that changing the order of belief elicitation and choice significantly affects subjects' beliefs. This simple change in task order does not accord with Nash equilibrium expectations. Are the subjects who guess after they choose simply winging it first and rationalizing later, or are the others simply winging their guesses first and then choosing according to something else later?

5 Discussion

Our results show, as is usually shown, that subjects deviate from NE predictions. They also show that these deviations are not simple, consistent, or easily explained: they depend on the specific setting and task. Our results also demonstrate that there are not 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.

The assumptions about human cognition that are part of the mythology of game theory, Nash equilibrium, and its refinements are at odds with what we know about actual human cognition. This is not a surprise, because the equilibrium concepts were not constructed based on how actual humans think, reason, or make decisions. Models that use false assumptions may not be problematic if our goal is to predict, rather than to understand, outcomes. However, Nash equilibrium and related models fail to predict behavior, which means we cannot resort to predictive success to justify the use of false assumptions.

We have shown that the protected core of game theory—the unrecognized cognitive model of non-cooperative game theory—fails repeatedly in hypothesis testing. We are not the first to say so, and the results are not shocking. Human beings for tens of thousands of years have been adept at moving through different settings and roles, exactly because they can adjust their demeanor, preferences, beliefs, and actions to suit their situations. Relative to all other species, they can turn on a dime, and this has made them astonishingly successful at inhabiting and constructing different forms of life. That people are not inflexibly fixed in their strategies does not mean that they are arbitrary and random. Rather than trying to fit people to a poor mythology, we should construct models that fit the reality of human behavior. We have not yet found a source—in neurobiology, computer science, evolution, or economics—from which this model can spring, like Athena from the head of Zeus, fully formed in convincing simplicity and power.

Acknowledgments. McCubbins acknowledges the support of the National Science Foundation under Grant Number 0905645. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Turner acknowledges the support of the Centre for Advanced Study at the Norwegian Academy of Science and Letters.

References

1. Camerer, Colin. Behavioral Game Theory. Russell Sage Foundation, New York, New York/Princeton University Press, Princeton, New Jersey. (2003)
2. Kahneman, Daniel and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." Econometrica, Vol. 47, No. 2. pp. 263-292. (1979)
3. Rabin, Matthew and Richard H. Thaler. "Anomalies: Risk Aversion," Journal of Economic Perspectives. vol. (1), pages 219-232, Winter. (2001)

4. Hoffman Elizabeth, McCabe Kevin, Shachat Keith and Smith Vernon. "Preferences, Property Rights, and Anonymity in Bargaining Games." Games and Economic Behavior 7(3): 346-380. (1994)
5. Rabin, Matthew. “Incorporating Fairness into Game Theory and Economics.” American Economic Review, LXXXIII. 1281-1302. (1993)
6. Costa-Gomes, M. A. and V. P. Crawford. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," American Economic Review. vol. 96(5), pp. 1737-1768. (2006)
7. Gigerenzer, Gerd and R. Selten, Reinhard Bounded Rationality: The Adaptive Toolbox. MIT Press (2002).
8. Simon, Herbert. "A Behavioral Model of Rational Choice", in Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting. New York: Wiley. (1957)
9. Stahl, D. and P. Wilson. “Experimental Evidence on Players’ Models of Other Players,” Journal of Economic Behavior and Organization, 25, 309–327. (1994)
10. Gigerenzer, Gerd. Adaptive thinking: Rationality in the real world. New York: Oxford University Press. (2000)
11. Gigerenzer, Gerd. Rationality for mortals: How people cope with uncertainty. New York: Oxford University Press. (2008)
12. Turner, Mark. 2009. “The Scope of Human Thought.” http://onthehuman.org/humannature/ .
13. McCubbins, Mathew D. and Mark Turner. “Going Cognitive: Tools for Rebuilding the Social Sciences.” In Sun, Ron, ed. Grounding Social Sciences in Cognitive Sciences. Cambridge MA: MIT Press. Chapter 14. (In press)
14. Fudenberg, D. and J. Tirole. Game Theory. MIT Press. (1991).
15. Nisan, N., T. Roughgarden, E. Tardos, and V. V. Vazirani. Algorithmic Game Theory. Cambridge University Press. (2007).
16. Lupia, A., A. S. Levine, and N. Zharinova. “Should Political Scientists Use the Self Confirming Equilibrium Concept? Benefits, Costs and an Application to the Jury Theorem.” Political Analysis 18:103-123. (2010).
17. Aumann, R., and A. Brandenberger. Epistemic conditions for Nash equilibrium. Econometrica 63, 1161–80. (1995)
18. Croson, R.. "Theories Of Commitment, Altruism And Reciprocity: Evidence From Linear Public Goods Games," Economic Inquiry, vol. 45(2), pages 199-216, 04 (2007)
19. Kuhlman, D. M., and Wimberley, D. L. “Expectations of choice behavior held by cooperators, competitors, and individualists across four classes of experimental game.” Journal of Personality and Social Psychology, 34, 69-81. (1976)
20. McKenzie, C. R. M., Mikkelsen, L. A. (2007). “A Bayesian view of covariation assessment.” Cognitive Psychology, 54, 33-61. (2007)
21. Berg, J. E., J. Dickhaut and K. McCabe. “Trust, Reciprocity, and Social History," Games and Economic Behavior, 10, 122-142 (1995)
22. Wolfers, J. and E. Zietzwitz. “Prediction Markets.” Journal of Economic Perspectives. American Economic Association, vol. 18(2), pages 107-126. (2004)
23. Sherrington, Charles Scott, Sir. Man on his Nature. New York: New American Library. (1964)
24. McCubbins, Mathew D., Mark Turner and Nicholas Weller. The Challenge of Flexible Intelligence for Models of Human Behavior. Technical Report of the Association for Advancement of Artificial Intelligence Spring Symposium on Game Theory for Security, Sustainability and Health. (2012)