Predictive Game Theory*

Drew Fudenberg

Harvard University

Abstract: Game theory is used in a variety of field inside and outside of social science. The standard methodology is to write down a description of the game and characterize its Nash or subgame perfect equilibria, but this is only sometimes a good approximation of observed behavior. The goal of predictive game theory is to develop models that better predict actual behavior in the field and in the lab. Core questions include: What determines people's behavior the first time they play an unfamiliar game? How do people update their play based on their observations? What sorts of "theories of mind," if any, are commonly used to guide play, and what do people think about the objective and rationality of their opponents? How do people think about games with a very large number of actions- what sort of "pruning" is involved? When will play resemble an equilibrium of the game, and which equilibrium will tend to emerge? Similarly, in a decentralized matching market, when will play converge to a stable outcome, and which one? To develop answers, researchers will need to combine insights from behavioral economics and psychology with formal modeling tools from economics and computer science.

Predictive Game Theory

1.   Research Agenda, importance, and context.

The standard methodology in applying game theory is methodology is to write down a description of the game and characterize its Nash or subgame perfect equilibria. This was a good starting point for game theoretic analysis, and has provided a number of qualitative insights It also yields a good approximation of observed behavior in some cases, but in many others it is either too vague to be useful or precise but at odds with how games are actually player. With the increased use of game theory in a variety of fields inside and outside of social science, it is time to go beyond equilibrium analysis to get more accurate predictions of behavior in the field and in the lab. There have already been some tentative steps towards this goal, from several different directions; the challenge is to go extend and perhaps unify these initiatives to build a coherent predictive theory.

A.   **Relaxing Equilibrium Analysis** A key component of this program is the further development of *adaptive* justification for equilibrium, which holds that equilibrium arises as the long-run outcome of a non-equilibrium process of learning or evolution. Existing work has focused on tractable learning rules that yield qualitative insights about long-run outcomes.

Researchers should now consider learning rules that more accurately describe how subjects update their play in light of their observations. One possibility is to take into account various cognitive limitations on learning that have been observed in decision problems, such as the use of coarse categories, errors in computing posterior probabilities, and so on. Also, the literature on adaptation and learning in extensive form games should move beyond the rational or almost-rational approach to off-path experimentation by considering other reasons that subjects might test the consequences of an apparently suboptimal action.

Other avenues for improvement is the addition of explicit models of the subjects "theories of mind"- their beliefs about how other subjects think about the game- and integrating "social preferences" into non-equilibrium learning theory: Even when monetary payoff functions are made public knowledge in the lab, some agent's objectives

may reflect such concerns as altruism, fairness, or spite, and the extent of such social preferences is rarely public information. How aware are subjects of these concerns? How does play correlate (within subject and across subject pools) with post-game survey questionnaires on social preferences?

In addition, researchers should begin to complement results on asymptotic behavior with results on the rate of convergence, and also with results that apply to laboratory settings, where subjects typically play ten, and at most fifty, repetitions of the game. In an extensive form game, even experienced players may not have learned how opponents respond to actions that have rarely if ever been used; as a result learning processes can converge to non-Nash outcomes such as those of self-confirming equilibria. Furthermore, in many cases of in the lab and in the field, agents do not have enough experience with the game to learn even the path of play, so that their initial beliefs and attitudes can play a large role in determining what is observed over the relevant horizon. This motivates a more careful and less agnostic treatment of the players' initial beliefs and attitudes.

This is related to the second key component of the program, the further development of models of cognitive hierarchies and level-k thinking. These models, which describe the outcome the first time people play an unfamiliar game, take as a primitive the players' beliefs about the play of unsophisticated "level-0" agents. Early work focused on simple matrix games, and supposed that level-0 agents give each action equal probability, but fitting these models to more complex games requires alternative ad-hoc modifications of level-0 play, and when all distributions over level-0 play are allowed the theory has very little predictive content. Thus, the cognitive hierarchy models should be complemented with an *a priori* method of determining level-0 play, one that takes social preferences into account. We also need a theory of how these beliefs are updated in light of observations and what the resulting play will be, which is especially important for applying make the technique useful for field data. Once again insights from behavioral psychology and economics should be brought to bear.

B. **Multiple Equilibria** Many games of interest have multiple equilibria, even when restricting to standard solution concepts, and allowing players to have incorrect off-path beliefs (as in self-confirming equilibrium) only makes the set of equilibria larger. Yet there is no general and empirically valid way of selecting between them. ere is a sizable theoretical literature that provides evolutionary/adaptive arguments for why cooperation should be observed in repeated games, but the existing theories are a poor match for the data from lab experiments: subjects do seem to cooperate when the gains to cooperation are sufficiently high, but do not cooperate in some settings that have cooperative equilibria. So when research question is to empirically characterize when cooperation occurs (varying payoff functions, what subjects observe about other subjects' play, etc.) and to then organize the findings in a way that makes testable predictions. There is also a sizable theoretical literature on "equilibrium refinements," and a literature using stochastic stability to select equilibria. The smaller experimental literature that has focused on the special cases of coordination games and signalling games; once again what is needed is an empirical characterization of behavior to serve as a constraint on theories of equilibrium selection.

C. Heuristics for Tree Pruning and Similarity

How do people simplify complex strategic interactions- what classes of strategies are viewed as equivalent and which ones are discarded? How do people extrapolate from past experience to one game to play in a "similar one, and what sorts of games are viewed as related? Ideas from computer science as well as psychology may be helpful here: computing the set of Nash equilibria of arbitrary large games is complex, but some classes of games have more parsimonious representations that allow polynomial-time complexity. These same ideas may permit more efficient estimation of behavior rules in complex economic environments, as the behavior rules are based on the agents' simplified models of the environment as opposed to the environment itself.

D. Matching Theory

Classic matching theory is based on the idea of a stable match, but stability is not a good approximation of the outcomes of laboratory experiments on decentralized matching except in extremely small markets with a unique stable outcome. When there are multiple stable outcomes, the analysis of decentralized markets closely parallels that of equilibrium analysis, and raises similar questions: when will a stable outcome will arise, and when it does, which one?

E. Empirical Validation

Work on predictive game theory should draw on lab and field data, and in many cases will be accompanied by explicit data analysis. Individual learning rules are notoriously hard to identify from laboratory data, so one focus will be the aggregate consequences of a population of agents using a distribution of rules. Another possibility is the use of exit surveys and in-game belief elicitations. A challenge in using field data is that the standard methodology imposes a form of subgame-perfect equilibrium as an identification condition to estimate model parameters. Recent work by Fershtman and Pakes relaxed this, allowing for players to maintain incorrect beliefs that are consistent with their observations. The challenges here are (1) to theoretically identify the sorts of equilibria that their algorithm tends to select, (2) test if the implicit equilibrium selection is stable over time and to changes in government policy, and (3) develop a way of testing if the equilibrium assumption is valid or if players have not even learned the path of play. A further challenge is to study non-equilibrium adaptation and learning on field data; this could be facilitated by running field experiments on the internet, either on "laboratory" sites or on commercial ones. Moreover, the current wave of internet-based field experiments would benefit from a grounding in the theory of non-equilibrium learning.

2.   Implications

This program will require the use and support of existing game theory labs, and may well justify the construction of new ones. It will also require graduate students who are trained in game theory, experimental methods, and econometrics; at present many of

the best theory students neglect these more applied domains. The program would also benefit from a more modern program for lab clusters than z-tree, with cleaner code and a more intuitive interface. Both the experimental and field components would benefit from improvements in computational game theory- this literature should continue to improve methods for computing Nash or subgame perfect equilibria in economically relevant games, but it should also take up the problems of computing and estimating equilibrium concepts that allow for incorrect off-path beliefs and/or cognitive errors, and of simulating and estimating non-equilibrium dynamics.

3.   Who is Doing Provocative Research?

The following very incomplete list is intended to give a sense of the scope of this agenda; it is not intended to be comprehensive or even a guide to the literature, and reflects the availability biases of the author. That said, Colin Camerer, Miguel Costas-Gomes, Vince Crawford, Tek Ho, Rosemarie Nagel, and Dale Stahl are leading the surge in work on cognitive hierarchies, and Armin Falk, Ernst Fehr and Simon Gachter are leading the experimental study of social preferences. Pedro Dal Bó, Anna Dreber, Guillaume Frechette, and Dave Rand are doing intriguing experimental work on cooperation in repeated games; Andrew Schotter has made provocative use of in-game belief elicitation. Ignacio Esponda, Philippe Jehiel, and David K. Levine are leaders in studying adaptive processes in extensive form games, and the sorts of non-Nash equilibrium outcomes that can persist even when players have a lot of experience with the game. Michel Benaïm, Josef Hofbauer, William Sandholm, and Sylvain Sorin are making important advances in the mathematics of dynamical systems and applying them to non-equilibrium dynamics. Many people are doing exciting work on cognitive limitations in decision problems, including Xavier Freixas, David Laibson, Sendhil Mullinaithan, and Matt Rabin, but so far little of this work has been applied to learning in games. Konstaninos Daskalakis and Tuomas Sandholm are exciting algorithmic game theorists with an interest in economic problems. Federico Echinique, Muriel Niederle, and Leeat Yariv are studying decentralized matching in the lab. Tim Salmon and Nathaniel Wilcox are pioneers in the econometrics of laboratory learning rules; Chaim

Predictive Game Theory

Fershtman and Ariel Pakes are developing estimation methods for field data that allow for incorrect off-path beliefs. Bernhard von Stengel is a leader of computational game theory, and Jeff Shamma is a pioneer in bringing techniques from the feedback-control literature to the study of learning in games.