Behavioural Game Theory: Predictive Models and Mechanisms

James R. Wright(✉)

University of British Columbia, Vancouver, BC, Canada
jrwright@cs.ubc.ca

1 Introduction

Classical economic models proceed from strong rationality assumptions which are known to be inaccurate (as no human is perfectly rational), but which are thought to reasonably approximate aggregate human behaviour. However, there is now a wealth of experimental evidence that shows that human agents frequently deviate from these models' predictions in a predictable, systematic way. Using this data, there is now an opportunity to model and predict human economic behaviour more accurately than ever before. More accurate predictions will enable the design of more effective multiagent mechanisms and policies, allowing for more efficient coordination of effort and allocation of resources.

Prediction (as distinct from description or explanation) is valuable in any setting where counterfactuals need to be evaluated, as where the impact of a policy needs to be determined before it is enacted. In this work, I am most interested in approaches that make explicit predictions about which actions a player will adopt, and that are grounded in human behaviour.

In a multiagent setting, perhaps the most standard game-theoretic assumption is that all participants will adopt Nash equilibrium strategies. However, experimental evidence shows that Nash equilibrium often fails to describe human strategic behaviour [e.g., Goeree and Holt, 2001]—even among professional game theorists [Becker et al., 205].

The relatively new field of *behavioural game theory* extends game-theoretic models to account for deviations from the standard models of behaviour, and proposes new models to account for human behaviour by taking account of human limitations [Camerer, 2003]. Researchers have developed many models of how humans behave in strategic situations based on experimental data. But this multitude of models presents a problem: which model should be used?

My thesis is that human behaviour can be predicted effectively in a wide range of settings by a single model that synthesizes known deviations from economic rationality. In particular, I claim that such a model can predict human behaviour better than the standard economic models. Economic mechanisms are currently designed under behavioural assumptions (i.e., full rationality) that are known to be unrealistic. A mechanism designed on the basis of a more accurate model of behaviour will be more able to achieve its goal, whether that goal is social welfare, revenue, or any other aim.

D. Barbosa and E. Milios (Eds.): Canadian AI 2015, LNAI 9091, pp. 356–359, 2015.
DOI: 10.1007/978-3-319-18356-5_35

2 Progress to Date

In my completed work I analyzed and evaluated several behavioural models for simultaneous-move games, eventually identifying a specific class of models (iterative models) as the state of the art. I then proposed and evaluated an extension that improves the prediction performance of any iterative model by better incorporating the behaviour of nonstrategic agents.

2.1 Model Comparisons [Wright and Leyton-Brown, 2010]

In my initial project, I explored the question of which of the quantal response equilibrium [McKelvey and Palfrey, 1995], level-k [Costa-Gomes et al., 2001; Nagel, 1995], cognitive hierarchy [Camerer et al., 2004], and quantal level-k [Stahl and Wilson, 1994] behavioural models was best suited to predicting out-of-sample human play of normal-form games. I also evaluated the standard game theoretic solution concept, Nash equilibrium.

Using a large set of experimental data drawn from the literature, I identified a single "best" model, quantal level-k, which performed best or nearly-best on each source dataset, plus a combined dataset. This is a striking result, as one might reasonably expect different models to perform well on different datasets.

2.2 Bayesian Parameter Analysis [Wright and Leyton-Brown, 2012]

In this work, I used a Bayesian approach to better understand the entire parameter space of two behavioural models: quantal level-k, the best-performing model identified in my previous work (Section 2.1); and cognitive hierarchy.

The parameter analysis identified several anomalies in the parameter distributions for quantal level-k, suggesting that a simpler model could be preferable. I identified a simpler, more predictive family of models based in part on the cognitive hierarchy concept. Based on a further parameter analysis of this family of models, I derived a three-parameter model, *QCH*, that predicts better than the five-parameter quantal level-k.

2.3 Level-0 Meta-Models [Wright and Leyton-Brown, 2014]

Iterative models such as QCH (Section 2.2) predict that agents reason iteratively about their opponents, building up from a specification of nonstrategic behaviour called level-0. The modeller is in principle free to choose any description of level-0 behaviour that makes sense for the given setting; however, in practice almost all existing work specifies this behaviour as a uniform distribution over actions. In most games it is not plausible that even nonstrategic agents would choose an action uniformly at random, nor that other agents would expect them to do so. In this work I considered "meta-models" of level-0 behaviour: models of the way in which level-0 agents construct a probability distribution over actions, given an arbitrary game. A linear weighting of features that can be computed from any normal form game achieved excellent performance across the board, yielding a combined model that unambiguously outperforms the previous state of the art.

3 Proposed Research

My proposed research will build upon my existing work on prediction in one-shot games, and extend it in two main directions: first, by leveraging machine learning techniques to improve the quality of the predictive models; second, by studying the implications of more accurate models of human behaviour for designing mechanisms and protocols.

3.1 Feature Discovery for Predicting Human Behaviour

Thus far, game properties that people might find *salient*—and hence be favoured by nonstrategic agents–are discovered primarily through introspection about particular examples, by asking oneself "How might I reason about playing this specific game?" *Deep learning* (see, e.g., [Bengio, 2009]) is a recent machine learning paradigm that has shown success in a wide range of (mostly signal processing) domains for automatically determining problem features. To my knowledge, deep learning has never before been applied in a game-theoretic context. I plan to study the adaptation of these techniques to the behavioural game theory domain to discover new representations of salient game characteristics.

I will begin work on this project by seeking to build highly predictive deep models. However, another important task will be integrating the discovered properties and insights into the framework of explicitly cognitive, lower-dimensional models, which I expect to be much easier to fit and use in applications.

3.2 Endogenous Levels

Most iterative models, including QCH, take the distribution of levels as a parameter. This implicitly assumes that the proportion of agents playing at a given level k will be identical regardless of the setting. This is unlikely to be true; rather, agents should be willing to perform more counterspeculation when it is easier or may plausibly yield greater rewards, and less otherwise.

In this phase, I will investigate ways of making the choice of level *endogenous* to the QCH model—that is, having the properties of the game or setting itself determine the distribution of levels, rather than having the distribution of levels be a parameter that must explicitly be learned.

3.3 Theoretical Implications for Mechanism Design

Standard mechanism design makes implicit assumptions about how agents will behave in response to incentives; typically that agents will play a Nash equilibrium. In this work, I will consider the implications for mechanism design of moving to a more accurate behavioural model. For example, what objectives are implementable according to an accurate model of behaviour, compared to those that can be implemented in equilibrium? What approximation guarantees are possible in terms of the properties of the agent population?

References

Becker, T., Carter, M., Naeve, J.: Experts playing the traveler's dilemma. Diskussionspapiere aus dem Institut fr Volkswirtschaftslehre der Universitt Hohenheim 252/2005, Department of Economics, University of Hohenheim, Germany (2005)

Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning **2**(1), 1–127 (2009)

Camerer, C., Ho, T., Chong, J.: A cognitive hierarchy model of games. QJE **119**(3), 861–898 (2004)

Camerer, C.F.: Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press (2003)

Costa-Gomes, M., Crawford, V., Broseta, B.: Cognition and behavior in normal-form games: An experimental study. Econometrica **69**(5), 1193–1235 (2001)

Goeree, J.K., Holt, C.A.: Ten little treasures of game theory and ten intuitive contradictions. AER **91**(5), 1402–1422 (2001)

McKelvey, R., Palfrey, T.: Quantal response equilibria for normal form games. GEB **10**(1), 6–38 (1995)

Nagel, R.: Unraveling in guessing games: An experimental study. AER **85**(5), 1313–1326 (1995)

Stahl, D., Wilson, P.: Experimental evidence on players' models of other players. JEBO **25**(3), 309–327 (1994)

Wright, J.R., Leyton-Brown, K.: Beyond equilibrium: predicting human behavior in normal-form games. In: Twenty-Fourth Conference of the Association for the Advancement of Artificial Intelligence, AAAI 2010, pp. 901–907 (2010)

Wright, J.R., Leyton-Brown, K.: Behavioral game-theoretic models: a Bayesian framework for parameter analysis. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, vol. 2, pp. 921–928 (2012)

Wright, J.R., Leyton-Brown, K.: Level-0 meta-models for predicting human behavior in games. In: Proceedings of the Fifteenth ACM Conference on Economics and Computation, pp. 857–874. ACM (2014)