Introduction to special issue on psychological game theory

What's your favourite article? Ours is "Psychological games and sequential rationality," written by John Geanakoplos, David Pearce and Ennio Stacchetti (1989) (GPS). The authors point out that classical game theory, where utilities depend only on actions, isn't a rich enough toolbox to capture a variety of psychological or social aspects of motivation. Yitzhak Gilboa and David Schmeidler (1988) reach similar insights. GPS address the issue via a new mathematical framework – psychological game theory – in which utilities are defined on beliefs (about actions and beliefs) as well as on actions. For an indication why this is special, recall GPS' opening example:

> Think of a two-person game in which only player 1 moves. Player 1 has two options: she can send player 2 flowers, or she can send chocolates. She knows that 2 likes either gift, but she enjoys surprising him. Consequently, if she thinks player 2 is expecting flowers (or that he thinks flowers more likely than chocolates), she sends chocolates, and vice versa. No equilibrium in pure strategies exists. In the unique mixed strategy equilibrium, player 1 sends each gift with equal probability. Note that in a traditional finite game with only one active player, there is always a pure strategy Nash equilibrium. That this is untrue in psychological games demonstrates the impossibility of analyzing such situations merely by modifying the payoffs associated with various outcomes […].

Matthew Rabin realized that GPS' framework is useful for modelling reciprocity, the inclination to be kind to those who are kind and to avenge those who are unkind. The idea of reciprocity is age-old. In economics it can be linked back to early 80s scholarship by George Akerlof on gift exchange in labour markets and by Werner Güth on ultimatum bargaining. Rabin had a key insight: kindness depends on intentions, hence beliefs, so when assessing beliefs about others' kindness one needs to consider beliefs about beliefs. This leads to a need for GPS' theory, and Rabin (1993) is a landmark reciprocity contribution based upon that recognition.

An important limitation of GPS' framework concerns which beliefs (about belief or actions) are allowed to influence utility. In the original formulation these are restricted to be initial beliefs, not beliefs that are updated based on how play proceeds. The issue is economically very important. For example, when modelling reciprocity, if players interact in a dynamic game then GPS's framework will not adequately capture how players' concerns for kindness and reciprocity should get updated depending on play. Rabin pointed to the issue, but did not address it formally. Martin Dufwenberg and Georg Kirchsteiger (D&K) (2004) took it from there, developing a reciprocity model for games with a dynamic structure in which updated beliefs influence players' kindness.

It is natural to wonder about the more general problem: Is it possible to meaningfully extend GPS' framework to allow infinite hierarchies of conditional beliefs to affect players' utilities? GPS themselves indicate that this would be worthwhile, although they left the topic for future work. Pierpaolo Battigalli and Dufwenberg (B&D) (2009) took it from there, developing the fuller framework. Crucial to their effort is the contribution by Battigalli and Marciano Siniscalchi (1999), which does not concern psychological games but which develops the mathematics needed to describe infinite hierarchies of conditional beliefs.

A host of interesting forms of motivations can be usefully explored using tools from psychological game theory. Apart from reciprocity, the sentiment that has gotten most attention is guilt aversion; see B&D (2007) for a general model and for references to related earlier work on specific games. And there are a host of other forms of motivation that either have not been explored at all, or which have been dealt with relatively little, at least via the framework of psychological game theory. This includes all emotions other than guilt, including disappointment, anger, joy, surprise, suspense, regret, hope, fear, and

anxiety. Another important class of sentiments involve what might be called concern with others' opinions, e.g. when a player likes it if others believe that he is ardent, honest, hard-working, or smart. Douglas Bernheim (1994) is a pioneering contribution of that sort, although he does not explicitly link his work to psychological game theory. (For such a link, see B&D 2009.)

Two additional research genres should be mentioned. First, psychological game theory may be employed to do applied economics, studying how belief-dependent motivation influences outcomes in various settings (e.g. marriages, banking, tax collection, MOUs, trade wars, etc.). Second, there is important room for empirical tests. Dufwenberg and Uri Gneezy (2000) ran the first lab experiment specifically set up to test an idea from psychological game theory (viz., guilt aversion), and their design included elicitation of a key-to-their-test second-order conditional beliefs. A natural extension concerns how communication between players may influence outcomes in psychological games, as Gary Charness and Dufwenberg (2006) were first to point out and test in the lab.

Awareness that psychological game theory is a useful framework is on the rise, yet we feel that too many scholars know too little about the exciting tools and topics at stake. We hope that this special issue will help counter that tendency. The included articles span a broad range of topics. We have grouped them into four categories: (i) basic framework; (ii) applied theory; (iii) experimental tests; and, (iv) surveys. We now offer brief comments on the articles, pointing out which category they fit with:

(i) Basic framework

Psychological game theory can seem rather complex. Pierpaolo Battigalli, Roberto Corrao and Martin Dufwenberg provide methodological guidance on how to model belief-dependent motivations and generate psychological games that can be explored in applied work. Relative to B&D (2009), a particular aspect is simplified (which beliefs may matter to utility) while others are generalized (e.g. the treatments of planning and of time) or modified (how beliefs are represented). The presentation also contains a broad and critical scrutiny of solution concepts, including rationalizability and self-confirming equilibrium. The article offers a gateway for scholars and students who are eager to learn how to work with psychological game theory.

(ii) Applied theory

We start this category with a pair of articles on how to model kindness and reciprocity. Andrea Isoni and Robert Sugden (I&S) argue that existing models, founded in psychological game theory, cannot explain the emergence of "trust" satisfactorily. At the core stand issues of how to define kindness. D&K, in their new paper, discuss the topic in depth, comparing Rabin's and their (2004) approach, articulating three reasons to favour the latter. They then provide a response to I&S, arguing that I&S have identified a flaw in D&K's (2004) definition which D&K (2019) repair.

The next pair concerns the economic implications of reciprocity, specifically for public good provision and contract design. Inducing truthful preference revelation is a classic problem in the public goods literature. Maria Kozlovskaya and Antonio Nicolò show that the standard pivot mechanism fails to solve this problem under reciprocity. Sequential implementation of the mechanism can however restore truth-telling. Luca Livio and Alessandro De Chiara consider whether a competitive or friendly workplace is more productive when workers are inclined to reciprocate. Using a principal-agent model they show that the optimal incentive scheme involves relative performance compensation if workers are only a little risk-averse and joint performance compensation if workers are sufficiently risk-averse.

Yet another pair of applied theory contributions focuses on a different motivation: guilt aversion. The papers consider two environments where guilt is a very natural concern: supply of public goods and communication. Alec Smith and Amrish Patel examine how guilt affects participation in public good provision. They characterize symmetric, mixed strategy equilibria and find that guilt can facilitate the existence of such equilibria and lead to non-monotonic comparative statics. Applying their model to existing experimental data on participation games they find a low but positive guilt sensitivity parameter. Kiryl Khalmetski studies the effect of guilt on the advice giving of experts. Analysing a model of strategic communication he finds that guilt avoidance may potentially explain why experts communicate in an evasive fashion. Given this, the receiver may prefer equilibria with fewer messages sent in equilibrium and to interact with senders who are less informed.

The final applied theory paper explores yet another belief-dependent motivation: a concern for social image. Alexander Sebald and Nick Vikander examine how a monopolist should set prices if consumers care about the beliefs that others hold about the product's popularity. They demonstrate that such a social image concern causes market distortions and they identify the implications for consumer demand and firm profits.

(iii) Experimental tests

Our five papers in this category illustrate how careful experimental design can be used to evaluate psychological game-theoretic predictions. The first paper should have been published time ago. There has been much interest in the last 10+

years in models of reference-dependent preferences in the style of prospect theory, but where the referent is given by beliefs which implies that psychological game theory is involved. Alec Smith ran a pioneering experiment, although others who did similar work got published before him. We are very glad to be able to now include his important study in our special issue.

Giuseppe Attanasi, Battigalli, Elena Manzoni and Rosemarie Nagel study guilt-aversion and reputation building in a finitely repeated trust game. They derive theoretical predictions involving incomplete information about belief-dependent preferences and design a novel experiment using belief disclosure to test their predictions.

The next two contributions provide new experimental insights on public goods provision. Sanjit Dhami, Mengxing Wei and Ali al-Nowaihi examine a public good game, taking into account that players might be motivated by guilt, reciprocity, as well as surprise-seeking. They report that all of these factors help explain contributions, and that guilt-aversion is more important than surprise-seeking. Stefano Caria and Marcel Fafchamps conduct an experiment in India to study how influential (or more connected) individuals contribute to public goods. They find that such individuals do not contribute significantly more than less influential individuals. However, they may be motivated to do so if others' expectations of co-players' contributions are disclosed, in line with some predictions based on psychological game theory.

Our final experimental contribution focuses on an economic problem particularly common in the developing world: embezzlement. Attanasi, Claire Rimbaud and Marie-Claire Villeval design a novel game to study how an intermediary's guilt affects embezzlement when donors' transfers to recipients go through intermediaries. They find that the intermediary's embezzlement is moderated if either the donor's or the recipient's expectation are believed to be high. This suggests that guilt towards another player can be triggered even when even when decisions have no direct monetary consequences for that player. B&D's (2007) guilt aversion theory presumes that guilt is felt only towards players that are hurt materially, so the indication is that this may be too restrictive.

(iv) Surveys

We conclude the issue with two surveys. Edward Cartwright focuses on the literature on guilt aversion in trust and dictator games. He reports that the evidence on balance seems consistent with the predictions of belief-based guilt aversion. Ofer Azar takes a bibliometric approach to examine the influence of psychological game theory. Specifically, he analyses citations of a selection of key papers on three topics: the general framework, reciprocity, and guilt aversion. The data shows that psychological game theory has influenced a number of disciplines beyond economics and that its influence continues to steadily increase.

The idea for a special issue was planted when the two of us organized the First and Second Workshops on Psychological Game Theory, at the University of Gothenburg in 2016 and at the University of East Anglia in 2017, each time generously sponsored by The Wallander Foundation. By the time that Giuseppe Attanasi organized the third instalment, in Soleto in Apulia in 2018, *JEBO*'s main editors, Dan Houser and Daniela Puzzello, had agreed to let us guest edit for them, and the call of papers was announced. We thank Dan and Daniela who have been tremendously supportive along the way. We received many submissions, and the process of evaluating these has been exciting. We are very grateful to the many scholars who helped us with refereeing, and we are proud of the fifteen articles that we accepted for publication. This fabulous collection provides great insight into the field of psychological game theory as it stands today. We hope many will enjoy reading the special issue, and then join the psychological game theory club.

Martin Dufwenberg
University of Arizona, United States
University of Gothenburg, Sweden
CESifo, Germany

Amrish Patel*
School of Economics, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

*Corresponding author.
E-mail addresses: martind@eller.arizona.edu (M. Dufwenberg), amrish.patel@uea.ac.uk (A. Patel)

References

Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Am. Econ. Rev. 97, 170–176.
Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. J. Econ. Theory 144, 1–35.
Battigalli, P., Siniscalchi, M., 1999. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. J. Econ. Theory 88, 188–230.

Bernheim, D.B., 1994. A theory of conformity. J. Polit. Econ. 102, 841–877.
Charness, G., Dufwenberg, M., 2006. Promises and partnerships. Econometrica 74, 1579–1601.
Dufwenberg, M., Gneezy, U., 2000. Measuring beliefs in an experimental lost wallet game. Games Econ. Behav. 30, 163–182.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games Econ. Behav. 47, 268–298.
Genakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games Econ. Behav. 1, 60–79.
Gilboa, I., Schmeidler, D., 1988. Information dependent games: can common sense be common knowledge? Econ. Lett. 27, 215–221.
Rabin, M., 1993. Incorporating fairness into game theory and economics. Am. Econ. Rev. 83, 1281–1302.