GAME THEORY: WHAT IT REVEALS ABOUT WHAT IS WRONG WITH MAINSTREAM ECONOMICS

Shaun P. Hargreaves Heap

Game theory has helped, more than anything perhaps, to define what is today's mainstream economics. This is because it is a theory of interactive decision making: that is, a theory of how people make decisions when they know that the consequences of their decisions depend on what other people also decide to do. In a complex economy where people rarely interact in textbook competitive markets, this fills what would otherwise be a rather large gap in economic theory. Further, since there are few social or political decisions that do not also fit this definition of interactive decision making, it has significantly contributed to the increasing dominance of mainstream economics as a way of thinking throughout the social sciences.

In this paper, I discuss two problems that have emerged and are now well recognised in Game Theory. Both relate to the status of the variously named, economic or rational choice or subjective expected utility maximising, model of how people decide what to do (I shall use rational choice in what follows). This is the model where it is assumed that people have preferences over outcomes and they act so as to satisfy best those preferences. Both the problems in game theory that I discuss speak to the inadequacy of this model. In the next section, I consider what must be further assumed to arrive at the Nash solution concept in game theory. In section 3, I discuss the problem of multiple Nash equilibria. In these sections, I also sketch two approaches in mainstream economics that have surfaced in relation to these problems: one makes us less rational and the other makes us more interesting than the rational choice model.

Hargreaves Heap, S.P. (2017)
'Game Theory: What It Reveals About What Is Wrong With Mainstream Economics'
Journal of Australian Political Economy
No. 80, pp. 74-87.

In the final section, I suggest that these two approaches to the weaknesses in Game Theory are an encouragement to engage in political economy.

Limits of Dominance Reasoning and the Nash equilibrium solution concept

Some games can be solved (that is provide guidance as to what a rational choice person should do) without any further assumption. The Prisoners Dilemma is, perhaps, the most famous example. It is illustrated in Figure 1. The pay-offs to each person from a pair of their actions are given by the numbers in each cell.

Figure 1

		Prisoner 2	
		C	D
Prisoner 1	C	3,3	0, 5
	D	5, 0	1, 1

In this interaction each person must choose between ‘cooperate’ (C) and ‘defect’ (D). In the original example ‘cooperate’ was cooperate with each other in denying the crime and ‘defect’ was defecting from such an agreement by confessing to the crime and implicating their fellow prisoner. It can be seen that the best course of action for each prisoner, no matter what they think the other prisoner will do, is to ‘defect’. If we assume that each player seeks only to do what is best for him or her self, then we conclude that the outcome of this game will be mutual defection. The game is important because it captures the essence of many fundamental interactions in social and economic life. For example, the multi person (or more than 2 person) version is the Public Goods game and this is thought to capture the key features of the interaction in a Hobbesian state of nature where each individual must decide whether to ‘arm’ (=D) or disarm (=C); as well as a myriad of actual public goods like defence, and aspects of education and healthcare. It also carries a

clear message. If rational individuals are left to make individual decisions in these settings, they will generate a sub-optimal result: mutual defection is worse for both than mutual cooperation. One way to avoid this outcome is through the creation of institutions of collective rather than individual decision making, like the State. The State removes the freedom to 'defect' and enforces mutual cooperation. This was Hobbes's argument.

There are, however, many games that cannot be solved by dominance reasoning. Figure 2 provides an example. I have put a * by the pay-off for each person that is the best a player can achieve given the action by the other player (in that cell). If Person 1 believes Person 2 is going to play 'a', then they will select either 'A' or 'B'. However, if 1 believes 2 is going to play 'b', then the best action is 'C'. Thus A's rational action depends critically on what he or she believes 2 will do. The difficulty now is that we cannot fix what 1 should believe about 2 simply by knowing that 2 is rational. This is because if 2 believes that 1 will play 'A' or 'C', then he or she chooses 'a', but if 2 believes that 1 will select 'B', then he or she prefers 'b'. In short, 1 could rationally choose A, B or C depending on the belief about whether 2 selects 'a' or 'b', but 2 can rationally choose 'a' or 'b' depending on his or her belief about 1.

Figure 2

		Person 2	
		a	b
Person 1	A	1*,4*	1, 3
	B	1*, 0	1, 5*
	C	0, 5*	2*, 3

This game does, however, have a unique Nash equilibrium. A Nash equilibrium consists of a strategy (i.e. an action) for each person such that each person's strategy is a best reply to the other people's strategies. In this case it is (A, a) as this is the only pay-off pair where each has a *. Further, it can be proved that every game has a Nash equilibrium (if not

in pure strategies then in mixed ones)[1]. Thus, in so far as the Nash equilibrium solution concept is consistent with rational choice rationality, then mainstream Game Theory is extraordinarily general. It can offer some advice about what rational choice people should do in *any* social and economic interaction that can be described as a game: that is, rational people will/should play a strategy that is in a Nash equilibrium with others. This is an amazing claim and it explains why Game Theory can claim to be a general theory: it has something to say about *any* strategic interaction. Perhaps, not unsurprisingly, in view of such a claim, the Nash equilibrium solution concept has attracted considerable critical scrutiny.

Since each action in a Nash equilibrium is the best for each player given the action of the others, the solution concept seems like a natural extension of rational choice reasoning to settings where dominance reasoning alone does not deliver a determinate outcome. In fact, it is now recognised that we must assume not only that each player is rational in the rational choice sense, but also that they each have common knowledge of rationality (that is, each knows that each is rational and knows that each knows that each knows that each is rational, and so on) and that each holds commonly aligned beliefs about what are rational inferences from any datum like a game. The additional assumptions are important. So long as a person knows through the common alignment of beliefs that there will be a unique inference about what rationality demands and if there is common knowledge of rationality in the rational choice model, then it follows that unique equilibrium must enjoy the property of being a Nash equilibrium. If it was not a Nash equilibrium, then someone would be expecting someone to do something that is not a best response to what others are doing and this contradicts the assumptions of rationality and common knowledge of rationality.

Although the assumption of common knowledge of rationality is obviously strong, it has been relaxed in various models of bounded rationality that have become part of the mainstream canon. For example, Level-k models allow for individuals to have different levels of strategic sophistication (see Crawford *et al* 2013). Level 0 players don't think strategically at all, they just choose. Level 1 players are strategic thinkers but they only take strategic thinking one stage by taking the best action

---

[1] Pure strategies consist of a single action; whereas mixed strategies involve a probabilistic mix of several pure strategies.

given that they think they are playing against level 0 players. Level 2 players have thought through two levels of strategic thinking and take the best action given that they think they are playing against level 1, and so on. How successful this is, is a matter of dispute (see Hargreaves Heap *et al* 2014). I want, however, to focus here on the other assumption that delivers the Nash solution concept: the common alignment of beliefs. This is also controversial and it has received less attention in mainstream game theory than common knowledge of rationality.

The assumption is usually justified by appealing to a famous argument by Aumann (1976): rational people 'cannot agree to disagree'. His point is roughly that, when people hold different beliefs, the fact of difference is a reason to adjust your own belief because there is something about the state of the world that another rational person discerns and which leads them to a different conclusion. It is a tempting line of argument, but it is controversial. It would, for instance, have surprised the arch-Rationalist philosopher, Kant, who thought that one of the key attributes of reason is that it knew its limits: there are certain things that we cannot know and this is what reason tells us. Perhaps more telling, though, it seems simply wrong as a matter of fact. Indeed it is bound to be so, so long as knowledge is advancing and we know, as a result, that, at any moment, our understanding of the world is imperfect.

This is perhaps obvious in the domain of scientific knowledge where it is plain that scientists hold different theories about the natural world, despite broadly sharing the same information set, without one or other being cast out as 'irrational'. This is because, when knowledge is imperfect, people can and will hold quite legitimately different views about the natural world; and whatever view they hold cannot be a consequence of reason alone. Reason only takes you so far. Most people believe for a variety of reasons that our understanding of the social world is bound to be less perfect than that of the natural world and so the likelihood of people disagreeing about what rationality requires in the social world seems, for this reason, even higher.

Another way of making this point, that connects it to other traditions in economics, is to say rationality is not very helpful when there is some uncertainty (as opposed to risk) in the social world. In such cases, something other than or as well as rationality, in the elaborated rational choice sense, explains our actions. There are many candidates for what this 'something' might be. It could be rules of thumb or heuristics and I

shall say more about this in the next section. The other candidate is emotions. Keynes (1936) famously thought many key economic decisions, like investment, both had a game theoretic interactive character (although this was not the terminology he used) and a temporal dimension that made them unavoidably uncertain. As a result, they were not occasions where rational calculation could gain a purchase. Instead, he argued, in the case of investment, that 'animal spirits' moved us to action. This is often the departure point for the Post Keynesian tradition in economics and although this tradition has been marginalised in the mainstream, I want to conclude this section by suggesting that one way of seeing what is wrong with mainstream game theory is that it slides over or ignores a key insight of Post Keynesian economics: the unavoidable limits of rational choice rationality when uncertainty (as opposed to risk) attaches to decision making.

Equilibrium selection where there are multiple Nash equilibria

The second weakness in mainstream Game Theory arises because many games have multiple Nash equilibria. Figure 3 gives the Stag hunt game. It is often thought to capture the interdependence of business investment decisions as well as the benefits of joint action that arise in many settings. There is a choice between hunting individually for a hare or collectively for a stag. There are two Nash equilibria (in pure strategies) in the game: [Hare, Hare] and [Stag, Stag]. Which Nash equilibrium strategy should rational agents select?

Figure 3

		Person 2	
		Stag	Hare
Person 1	Stag	20,20	0,10
	Hare	10, 0	10, 10

The Nash refinement project is largely a response to this problem. The project expands the set of considerations that rational choice agents use when making a decision. Several such auxiliary considerations have been proposed. For example, in this game, an additional attention to risk dominance would give a reason for selecting [Hare, Hare] because this yields a certain outcome (10) whatever the other person does. An attention to pay-off dominance, however, would give a reason for selecting [Stag, Stag] because this gives better pay-offs for both (20). The difficulty with both auxiliary motivational guides is that neither is clearly connected to the axioms of rational choice. So, there is no obvious reason for rational choice agents to rely on one rather than another. In short, the problem of selection remains. This is a general feature of the refinement project.

An alternative approach is to subtract from the rationality of agents by allowing that their decisions are made using rules of thumb or heuristics of one kind or another. This approach has typically drawn on the insights of psychology where the framing of a decision can, for example, either be in terms of gains or losses and it is known that risk taking behaviour is more likely to be triggered in the loss than in the gain domain (see Kahneman and Tversky 1979). The framing might, therefore, be important for whether risk dominance guides agents in Hawk-Dove. Schelling's (1960) introduction of the concept of salience is a similar analytic move. He suggest that the 'labels' ('hare' or 'stag' in this instance) attached to different actions render some more salient than others even though they may enjoy the same status as being Nash equilibrium strategies.[2] This is not a matter of reflection. It just happens and he introduced some very suggestive classroom experiments to demonstrate this. In these cases, it matters that the rules or heuristics are shared: otherwise they would not function to coordinate actions on one of the Nash equilibria. So, it is not enough that we use rules, they have to be shared and an obvious question concerns how these shared rules arise. This is where evolutionary game theory offers an account (see Sugden 1986).

If we follow the evolutionary approach, the rules we use are the product of chance in a history where extraneous bits of information (from a strict game theoretic point of view), like the height, age or gender of the

---

[2] The labels are irrelevant from a strict game theoretic point of view because, for rational choice agents, it is the pay-offs that matter and not the name of the action leading to them.

decision makers are used in decision making. To see how this works, imagine a crossroads-like interaction where one party must give way or there is a smash. When people start to use a piece of extraneous information like age and gender to condition whether they give way or not, then a rule like 'give way to men in mixed encounters or the older person in same sex encounters' could easily spread. This is because the moment some people use this rule, it will typically pay for others to do the same just because this solves the coordination (equilibrium selection) problem when they interact. No one need buy into the rule in the sense that they believe the old and men deserve to have priority: the key is that it is a practice that avoids smashes. Of course, people may believe in the substance of the rule, but what explains the spread of the rule and its prevalence in society is not its substance. Instead it is the rule's material coordinating function that explains this.[3] We are less rational and more historical, one might say, than the rational choice model suggests.

In some games, like negotiations over how to divide a 'cake', where there are multiple Nash equilibria (e.g. 50-50, 30-70, 60-40 and so on), there is another possible explanation of how people decide. (This is the approach that I referred to earlier as making us 'more interesting'.) They use a principle of justice. People share an idea about what is fair and this guides them to one outcome rather than another. This can be made consistent with the rational choice model by introducing the concept of a social preference. These are preferences that individuals have over the outcomes of the game both for themselves and for others. Thus getting 50% of the pie may be rated in terms of pay-offs more highly than getting 60% because getting 60% introduces a negative social preference pay-off to the individual when he or she deems that this is unfair. In this way, a game with multiple Nash equilibria in terms of purely selfish pay-offs can be transformed into one where there is only one Nash equilibrium when pay-offs include the return to people's social preferences. Typically, it will matter that the social preferences are shared by all the players in some degree so that they select the same equilibrium, but, this aside, the approach appears elegantly to keep faith

---

3 Marx famously makes a similar, if different in detail, functional argument at the beginning of in the Preface to the *Contribution to the Critique of Political Economy*: 'The mode of production of material life conditions the general process of social, political and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness'.

with the rational choice model while potentially solving the equilibrium selection problem in such games. This impression, however, can be deceptive for several, related reasons.

First, it is not clear that all types of principles of justice can be represented in this way by a social preference, particularly those that are keyed into the way the options and outcomes arise not just the outcomes themselves. One might feel differently about how to divide the cake depending, for example, on whether it was jointly baked by the players or had first been stolen from one of the players who baked it alone.

Second, there is considerable experimental evidence that people's social preferences change with social and institutional context. For example, people appear to be more other-regarding in non-market settings than in market ones (see Bowles and Polania Reyes 2012), but appear less influenced by status differences in market than non-market ones (see Hargreaves Heap *et al* 2013). Indeed both findings may reflect the point above: i.e. that people value the process determining the outcomes and not just the outcomes themselves. This is worrying because people's preferences are a primitive (that is, they are taken as 'given' and form the starting point for the analysis) in game theory and in mainstream economics more generally.

One way of seeing why preference change/variability is a problem is to think of how mainstream economics explains the origins of institutions (sometimes called the 'new institutionalism' following the work of Coase 1937 and Williamson 1975). They are the deposit of some antecedent decision problem. The Hobbesian explanation of the State, that I sketched earlier, is an obvious illustration: rational choice agents contract to create the State as an improvement over the sub-optimal outcomes that would otherwise occur. Likewise, the choice between organizing an exchange via the institution of a market or within the institution of the firm depends on the relevant respective transaction costs under each institutional arrangement. People use the institution with the lowest transaction costs. The institution selected is, thus, the one that best satisfies the preferences of the parties to the exchange. It is the most efficient. But here is the rub. It is assumed that people's preferences are the same whatever the institutional arrangement. This is what allows for institutions to be determined on grounds of efficiency. If people's preferences change with the choice of institution, this no longer holds. The institution is not simply a device to satisfy best some antecedent

preferences: the choice of institution involves in part a decision over what preferences to have. In other words, it becomes in part a choice over who to be; and this requires a way of thinking and deciding that does not turn on preference and preference satisfaction alone.

Finally, once people have social preferences, there is an obvious question that needs answering. How do people come to hold a social preference that is in important respects shared with others? Economics would seem to be drawn, in short, into an area that it does not naturally inhabit because the terrain belongs to sociology and politics: preference formation. As Duesenberry (1960) once quipped 'Economics is all about how people make choices; sociology is about how they don't have any choices to make'.

Summary: an encouragement to political economy

I have identified two problems in mainstream Game Theory. The most glaring is that of multiple equilibria. I have illustrated it with two important and interesting games (Hawk-Dove and Bargaining over shares of a 'cake'). It is also a pervasive problem because we know that once any game is indefinitely repeated, there are multiple Nash equilibria even when there is a unique Nash equilibrium in the stage game. This is the Folk Theorem. This means that the rational choice model frequently underdetermines what action a rational agent will take in economic and social interactions. The problem for the rational choice model is even deeper than not knowing which Nash equilibrium to select because, and this is the first problem I discussed, rational choice agents have to hold additional beliefs about this rationality being shared and coming to common conclusions if they are to be persuaded to select Nash equilibrium strategies in the first place. In short, the rational choice model is too slight a model of individual decision making to explain what people do in a variety of economic and social interactions.

I want to suggest in conclusion that these weaknesses are directly connected to the critique of mainstream economics that comes from political economy. In other words, the weaknesses are important for political economy. I will do this in two ways.

First, I want to make explicit what I have already alluded to when hinting that the internal responses to these weaknesses sometimes look surprisingly like the rediscovery of an 'old' Post Keynesian or Marxian

or sociological insight. The point here is that the two internal responses to these weaknesses, the introduction of rule-following bounded rationality and social preferences, are directly related to the political economy critique. Of course, there are many elements to this critique. But one relates to the reductive, equilibrium nature of mainstream economics. Once people have to share rules for making decisions to solve the problem of equilibrium selection, the analysis has to become non-reductive and historical and, with these features, less in thrall to the concept of equilibrium as the fulcrum for theoretical insight.

Another element of the critique, related to the reductive concern, is the presumed primitive (*i.e.* foundation) of individual preferences in mainstream economics and the way that this skews normative economics towards the discussion of preference satisfaction and away from the character of the preferences. Once people are motivated by social preferences and their existence/nature depends on institutional context or the policies in play, then it is no longer possible to judge institutions or policies simply in terms of preference satisfaction. Antecedent preferences are not given and the choice of institution turns in part on what preferences to have. In short, normative economics cannot avoid thinking about who we want to be in this sense; and this will require a significant engagement with philosophy and political theory. In particular, since social preferences often encode ideas about justice, normative economics is not going to be able to avoid or relegate concerns with justice in the way that the mainstream habitually does or has.

Second, I want to illustrate why an understanding of these weaknesses in the mainstream might be important for the practice of political economy: that is, for the actual analysis of contemporary events. It turns on the idea that today's neoliberalism is, in important respects, a child of mainstream economic thinking. If one accepts this idea, then it is reasonable to suppose that some of the 'problems' in contemporary neoliberalism will be refracted versions of the problems within mainstream economic theory. I think this is, indeed, the case. To make this case, I offer two examples of where responses to the weaknesses I have sketched have plausibly been in play in significant recent developments.

The first is the behaviour leading up to the 2008 financial crisis. It plausibly involved people in a Hyman Minsky fashion, becoming more optimistic about the security provided by housing collateral as house

prices rose. This led them to advance further housing loans which fuelled demand for houses and further house price rises. The reverse process took hold once house prices started to fall: pessimism grew as prices fell and credit markets froze. The point is that this is good example of where people in an uncertain world fall back on rules; and a key feature of such rules in financial markets is that they have to be shared. This is because the problem facing financial market participants is always that they need to coordinate their expectations about the future with others. The only alternative is to have very deep pockets to back a judgment that goes against the conventional wisdom and, even then, you will do better to follow the conventional wisdom until just before it collapses. The rule governing expectations that allows optimism to grow as house prices rise is just that: it is a rule that works in what is an uncertain world, at least for a while, because, when shared, it coordinates the behaviour of participants in financial markets.

The second example relates to the 'populist' electoral turn in several European countries and the US. One account of this begins with the fact that many low income households have experienced either stagnant or falling living standards for a considerable period of time in these countries. The (neoliberal) claim regarding the efficiency of the social order increasingly sounded hollow for this group. What is interesting is that their votes did not swing towards traditional parties offering redistribution. Had they done so, it would have been politics as usual with people voting in their material self interests. Instead, they voted for political outsiders whose programmes offered small prospect of redistribution in their favour. They were not voting on grounds of self interest. But they were suddenly motivated to vote; and typically in larger numbers than in previous elections. The question is: why? An obvious and common explanation is that they were protesting against the legitimacy of political system that had neglected their interests for so long. That is why they voted for outsiders. Another way of saying this is that they were motivated to act on social preferences regarding the justice of the economic system. In other words, you need a category of social preference (just as you do when solving some equilibrium selection problems in game theory) to understand populism.

To summarise, the argument is that a more complicated model of individual action than the standard rational choice one is required to solve problems in game theory and the kinds of amendments that might be made for this purpose (the introduction of rule governed and social

preference driven behaviour) are also what we need to understand key recent developments. In this respect, I'm inclined to end on a slightly self-congratulatory note because Yanis Varoufakis and I concluded our text on Game Theory (2004) in a similar vein without knowing how events like these would unfold:

> The problems we have identified with game theory resurface as timely warnings of the difficulties any society is liable to face if it thinks of itself only in terms of liberal individualism (p.303).

I can't do that, however. Our argument concerned 'liberal individualism'. We did not use the term 'neoliberalism', although this was the target we had in mind. In retrospect, we should have been more careful to distinguish this variety of liberalism from its classical counterpart, coming from, among others, J.S. Mill. The distinction is important because classical liberalism always argued that an individual's commitment to liberalism should also entail a commitment to being constrained in the exercise of liberal freedoms. Thus classical liberals are never just preference satisfiers. In so far as they do follow preferences, then they are also constrained by internal (as well as, in practice, external) rules regarding how preferences can be satisfied. What these rules might be is/was, of course, a matter of dispute. However, there is one principle that is foundational for classical liberalism: people are free to do whatever takes their fancy but only so far as their actions do not directly harm others. This is the no-harm principle and, although its application can be tricky given the potential elastic understandings of 'harm', it illustrates my argument here in a rather telling way for game theory (see Amadae 2015 for a detailed development of this argument).

I mention this now because it illustrates the general point of the argument here: that is, economics needs a model of individual action that does not turn simply on preferences and preference satisfaction. It is also rather telling because, at least on one understanding of the no-harm principle, this classical liberal addition to the preference satisfaction model would preclude rational agents defecting in the prisoner's dilemma. To do so, while an exercise of individual freedom in support of preference satisfaction, would be directly to do harm to another: i.e. make them worse-off. This runs against the no-harm principle that guides classical liberals. So it could not be right.

When was the last time you saw this classical liberal version of the argument about how rational people should choose in the Prisoner's

dilemma in a text on Game Theory? Your answer, I suspect, tells you in a nutshell everything about what is wrong with mainstream game theory.

Shaun P. Hargreaves Heap is Professor of Political Economy at King's College, London.

s.hargreavesheap@kcl.ac.uk

References

Amadae, S. (2015) *Prisoners of Reason: Game Theory and Neoliberal Political Economy*. New York: Cambridge University Press.

Aumann, R. (1976) Agreeing to Disagree, *The Annals of Statistics*, 4(6), 1236-1239

Bowles, S. and S. Polania Reyes 2012, 'Economic Incentives and Social preferences: substitutes or complements' *Journal of Economic Literature* 50(2), 368-425.

Coase, R. (1937) The Nature of the firm, *Economica*, 4(16), 386-405.

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri (2013) 'Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications'. *Journal of Economic Literature*, 51(1): 5-62.

Duesenberry, J. (1960). Comment on Gary Becker's 'An Economic Analysis of Fertility' in *Demographic and Economic Change in Developing Countries*. New York: National Bureau of Economic Research.

Hargreaves Heap, S., Tan, J. and D. J. Zizzo (2013) 'Trust, inequality and the market', *Theory and Decision*, 74, 311-333.

Hargreaves Heap, S., Rojo Arjona, D. and R. Sugden. (2014) 'How portable is level-0 behavior? A test of Level-k theory in games with non-neutral frames', *Econometrica*, 82(3), 1133-1151.

Hargreaves Heap, S. and Y. Varoufakis. (2004) *Game Theory: A critical text*. London: Routledge.

Kahneman, D. and A. Tversky. (1979) 'Prospect Theory: an analysis of decision under risk', *Econometrica*, 47(2), 263-292.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. London: Macmillan

Schelling, T. (1960) *The Strategy of Conflict.* Cambridge: Harvard University Press.

Sugden, R. (1986) *The Economics of Rights, Cooperation and Welfare*. Oxford: Basil Blackwell.

Williamson, O. (1975). *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*. New York: Free Press.