Lecture 07 : Philosophical Issues in Behavioural Science

Stephen A. Butterfill
< s.butterfill@warwick.ac.uk >

Tuesday, 1st March 2022

Contents

1. Introduction

To understand rational behaviour in social interactions we turn our attention to game theory.

Our aim in studying game theory is to answer this question:

> When two or more agents interact, so that which outcome one agent's choice brings about depends on how another chooses, how do their preferences guide their choices?

We will also investigate whether reflection on game theory provides a challenge to the leading philosophers' accounts of joint action.

This lecture depends on you having studied some sections from a previous lecture:

- *Expected Utility* in Lecture 06
- *What Are Preferences?* in Lecture 06
- *Bratman on Shared Intentional Action* in Lecture 04

For the minimum course of study, consider only these sections:

- *Game Theory Introduction* (section §2)
- *Nash Equilibrium* (section §3)

None of this week's material is required for the assignment on decision theory.

2. Game Theory Introduction

The bare minimum you need to know about game theory for the purposes of this course.

2.1. What Is a Game?

Different researchers offer different statements. Games are various characterised as interactions, decriptions of interactions and situations:

> A game is 'any interaction between agents that is governed by a set of rules specifying the possible moves for each participant and a set of outcomes for each possible combination of moves' (Hargreaves-Heap & Varoufakis 2004, p. 3)

> 'A game is a description of strategic interaction that includes the constraints on the actions that the players can take and the play-

> ers' interests, but does not specify the actions that the players do take' (Osborne & Rubinstein 1994, p. 2).
>
> 'All situations in which at least one agent can only act to maximize his utility through anticipating (either consciously, or just implicitly in his behavior) the responses to his actions by one or more other agents is called a game' (Ross 2018).

Although the different characterisations of games are probably not strictly equivalent, the differences are unlikley to matter for our purposes.

We will focus on noncooperative games which are one-off events (so not repeated).

2.2. Books

There are many different game theory text books you could use. Tadelis (2013) and Osborne & Rubinstein (1994) are relatively concise and formal. Hargreaves-Heap & Varoufakis (2004) is more chatty and probably easier to get started with, but my impression is that it is sometimes difficult to get a clear sense of what game theory is from this book. Dixit et al. (2014) is a beautifully written and very clear book that takes things quite slowly; any of the five editions in the library will be fine, but select a later edition if you have the choice.

2.3. Why Study Game Theory and Its Limits?

Our overall concern is with understanding joint action in particular and social interaction more generally (see *Introduction: Why Investigate Philosophical Issues in Behavioural Science?* in Lecture 01). Many researchers imply that game theory is relevant to this concern:

> 'we treat game theory not as a branch of mathematics but as a social science whose aim is to understand the behavior of interacting decision-makers' (Osborne & Rubinstein 1994, p. 2; compare Dixit et al. 2014, pp. 36–7).

and:

> 'game theory is the most important and useful tool in the analyst's kit whenever she confronts situations in which what counts as one agent's best action (for her) depends on expectations about what one or more other agents will do, and what counts as their best actions (for them) similarly depend on expectations about her' (Ross 2018).

Notably, even critics of game theory suggest that it is useful for understanding social interaction:

> 'understanding why game theory does not, in the end, constitute the science of society (even though it comes close) is terribly important in understanding the nature and complexity of social processes' (Hargreaves-Heap & Varoufakis 2004, p. 3)

For sources on applications of game theory to understanding law, conflict and foraging (among other things), see *Consequences and Applications of Game Theory* (section §4).

3. Nash Equilibrium

A nash equilibrium for a game is a set of actions from which no agent can unilaterally profitably deviate (Osborne & Rubinstein 1994, p. 14).

Game theory is supposed to explain why things happen:

> 'Many events and outcomes prompt us to ask: Why did that happen? [...] For example, cutthroat competition in business is the result of the rivals being trapped in a prisoners' dilemma' (Dixit et al. 2014, p. 36).

This section introduces two notions that are involved in giving such explanations, dominance and Nash equilibrium.

If you understand these notions and can apply them, you can do game theory.

3.1. Nash Equilibrium

A Nash equilibrium for a game is a set of actions (sometimes called a 'strategy') from which no agent can unilaterally profitably deviate.

Why *equilibrium*?:

> 'equilibrium [...] simply means that each player is using the strategy that is the best response to the strategies of the other players' (Dixit et al. 2014, p. 32–3)

Although not covered in this section, there is some interesting research on other ways of specifying a 'best response' (Misyak & Chater 2014a,b). Why might you want to do so? Potential motives arise in *Consequences and Applications of Game Theory* (section §4) and *What Is Team Reasoning?* (section §7).

4. Consequences and Applications of Game Theory

4.1. Successes and Failures

Problems for applications of game theory are easy to find Hargreaves-Heap & Varoufakis 2004 is particularly full of them, but any recent-ish textbook will cover some.

What's puzzling about game theory is that, despite the problems, there are many cases where it is successfully used to explain things.

This section introduces one case where game theory has been successfully used to explain behaviour (Sinervo & Lively 1996). There are many others, including:

- in law: inequality, culture and power (McAdams 2008)
- network security (Roy et al. 2010)
- evolution of social contract (Skyrms 2000)
- distribution of water resources (Madani 2010)
- the tragedy of the commons Tadelis (2013, §5.2.2)
- foraging behaviours (Hansen 1986)

If studying game theory, it would be a good idea to consider how it has been applied in a domain of interest to you.[1]

5. Index of Games

Lecturer: Stephen A. Butterfill

If you are looking for a particular game like Hi-Lo, the Prisoner's Dilemma or Hawk-Dove, it should feature in these slides.

6. Why Investigate Team Reasoning?

There are at least three motives for us to investigate team reasoning. It provides a development of game theory which arguably better captures the notion of rational choice in many ordinary social interactions. It promises to provide an explanation of how there could be aggregate subjects. And it might provide an account of shared intention.

This section introduces some motives for investigating team reasoning. Another section, What Is Team Reasoning? *(section §7), explains what team reasoning*

[1] I am not particularly recommending the sources cited here. Please share with me any good sources you find.

is and the justification for supposing that it exists.

6.1. Prerequisites

This section depends on you having studied some other sections:

- *Game Theory Introduction* (section §2)
- *Nash Equilibrium* (section §3)

6.2. Applications of Team Reasoning

Team reasoning can be drawn on in attempting, perhaps not always successfully, to provide:

- an account of rational decision which differs from plain vanilla game theory on what is rational in many ordinary social interactions which have the structure of games like the Prisoner's Dilemma[2] and Hi-Lo[3] (Bacharach 2006; Sugden 2000)
- an alternative to *Bratman on Shared Intentional Action* in Lecture 04 (Gold & Sugden 2007; Pacherie 2013)
- an explanation of how there could be aggregate subjects.

7. What Is Team Reasoning?

'You and another person have to choose whether to click on A or B. If you both click on A you will both receive £100, if you both click on B you will both receive £1, and if you click on different letters you will receive nothing. What should you do?' (Bacharach 2006, p. 35) Team reasoning is a game-theoretic attempt to explain what makes your both choosing A rational. But what is team reasoning?

7.1. Prerequisites

This section depends on you having studied some other sections:

- *Game Theory Introduction* (section §2)
- *Nash Equilibrium* (section §3)

[2] These games are specified in the *Index of Games* (section §5)

[3] These games are specified in the *Index of Games* (section §5)

7.2. Aim

This section provides an informal explanation of team reasoning starting from Bacharach's initial characterisation:

> 'somebody team reasons if she works out the best possible feasible combination of actions for all the members of her team, then does her part in it.' (Bacharach 2006, p. 121)

7.3. Alternative Approach

Although not covered in these lectures, Misyak & Chater (2014a)'s proposal about virtual bargaining also looks like a promising development of game theory.

8. Question Session 07

These are the slides I prepared for the question session. In the end we had a small-group discussion about just part of this. Because the event turned into a discussion, there is no recording. You are of course welcome to ask questions.

8.1. Mixed Strategies (Alex' Question)

Optional. You do not need to know about mixed strategies for this course.

In many situations you might want to vary how you act rather than always acting in the same way. You may benefit from making your actions unpredictable to others. The game-theoretic notion of a mixed strategy is intended to capture this.

To illustrate, suppose you are playing the game rock, paper, scissors (see *Index of Games* (section §5)). It would not be good if your opponent could predict your action. How might you play? One possibility would be to try to pick each action with equal probability, but in a way that was unpredictable. To illustrate, you might toss a three-sided dice and decide what to do based on which side it lands on.

In a mixed strategy, one or more players does not simply select an action to perform but rather assigns weights to the different actions and then selects one at random in such a way that the probability of selecting an action matches the weight assigned to it.

To illustrate, in *hawk-dove* (see *Index of Games* (section §5)), Gangster Y might decide to play *stay home* with probability 0.75 and *attack* with probability 0.25.

They expected payoff from a mixed strategy is obtained by calculating the expected payoff for each action and multiplying it by the probability that the action will be performed if the mixed strategy is implemented. (See Tadelis (2013, §6.1.4) for details.)

The notion of a Nash equilibrium can be extended to mixed strategies:

> 'Nash equilibrium is defined as a list of mixed strategies, one for each player, such that the choice of each is her best choice, in the sense of yielding the highest expected payoff for her, given the mixed strategies of the others.' (Dixit et al. 2014, p. 216; see Osborne & Rubinstein 1994, definition §32.3 for a more formal statement)

Glossary

aggregate subject A subject whose proper parts are themselves subjects. A paradigm example would be a Portuguese man o' war (Physalia physalis), which is an animal that can swim and eat and whose swimming and eating is not simply a matter of the swimming or eating of its constituent animals. Distinct from, but sometimes confused with, a plural subject. 5, 6, 9

decision theory I use 'decision theory' for the theory elaborated by Jeffrey (1983). Variants are variously called 'expected utility theory' (Hargreaves-Heap & Varoufakis 2004), 'revealed preference theory' (Sen 1973) and 'the theory of rational choice' (Sugden 1991). As the differences between variants are not important for our purposes, the term can be used for any of core formal parts of the standard approaches based on Ramsey (1931) and Savage (1972). 2

dominance An action (or strategy) *strictly dominates* another if it ensures better outcomes for its player no matter what other players choose. (See also weak dominance.) 4

game theory This term is used for any version of the theory based on the ideas of von Neumann et al. (1953) and presented in any of the standard textbooks including. Hargreaves-Heap & Varoufakis (2004); Osborne & Rubinstein (1994); Tadelis (2013); Rasmusen (2007). 2, 5, 6, 9

mixed strategy In game theory, a *mixed strategy* for a player is a probability distribution over the actions available to the player. 7

Nash equilibrium a profile of actions (sometimes called a 'strategy') from which no agent can unilaterally profitably deviate. 4, 8

noncooperative game 'Games in which joint-action agreements are enforceable are called *cooperative* games; those in which such enforcement is not possible, and individual participants must be allowed to act in their own interests, are called *noncooperative* games' (Dixit et al. 2014, p. 26). 3

plural subject Some subjects who are collectively the subject of an intention or other attitude. If there is one token intention that is both my intention and your intention and no one else's intention, then we are the plural subject of that intention. (The intention is therefore shared in the same sense that, if we were siblings, we would share a parent.) Distinct from, but sometimes confused with, an aggregate subject. 8

shared intention An attitude that stands to joint action as ordinary, individual intention stands to ordinary, individual action. It is hard to find consensus on what shared intention is, but most agree that it is neither shared nor intention. (Variously called 'collective', 'we-' and 'joint' intention.) 5

strict dominance In game theory, one action *strictly dominates* another action if the first action guarantees its player higher payoffs than the second action regardless of what other players choose to do. (See Definition 59.2 in Osborne & Rubinstein 1994, p. 59 for a more general definition.) 9

team reasoning 'somebody team reasons if she works out the best possible feasible combination of actions for all the members of her team, then does her part in it' (Bacharach 2006, p. 121). 5, 7

weak dominance In game theory, one action *weakly dominates* another action if the first action guarantees its player payoffs at least as good as the other action and potentially better than it regardless of what other players choose to do. (Contrast strict dominance.) 8

References

Bacharach, M. (2006). *Beyond Individual Choice*. Princeton: Princeton University Press.

Dixit, A., Skeath, S., & Reiley, D. (2014). *Games of Strategy*. New York: W. W. Norton and Company.

Gold, N. & Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy*, *104*(3), 109–137.

Hansen, A. J. (1986). Fighting Behavior in Bald Eagles: A Test of Game Theory. *Ecology*, *67*(3), 787–797.

Hargreaves-Heap, S. & Varoufakis, Y. (2004). *Game theory: a critical introduction*. London: Routledge.

Jeffrey, R. C. (1983). *The Logic of Decision, second edition*. Chicago: University of Chicago Press.

Madani, K. (2010). Game theory and water resources. *Journal of Hydrology*, *381*(3), 225–238.

McAdams, R. H. (2008). Beyond the Prisoners' Dilemma: Coordination, Game Theory, and Law. *Southern California Law Review*, *82*(2), 209–258.

Misyak, J. B. & Chater, N. (2014a). Virtual bargaining: A theory of social decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130487.

Misyak, J. B. & Chater, N. (2014b). Virtual bargaining: A theory of social decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130487.

Osborne, M. J. & Rubinstein, A. (1994). *A course in game theory*. MIT press.

Pacherie, E. (2013). Intentional joint agency: shared intention lite. *Synthese*, *190*(10), 1817–1839.

Ramsey, F. (1931). Truth and probability. In R. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays*. London: Routledge.

Rasmusen, E. (2007). *Games and Information: An Introduction to Game Theory* (4th ed ed.). Malden, MA ; Oxford: Blackwell Pub.

Ross, D. (2018). Game Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.

Roy, S., Ellis, C., Shiva, S., Dasgupta, D., Shandilya, V., & Wu, Q. (2010). A Survey of Game Theory as Applied to Network Security. In *2010 43rd Hawaii International Conference on System Sciences*, (pp. 1–10).

Savage, L. J. (1972). *The Foundations of Statistics* (2nd rev. ed ed.). New York: Dover Publications.

Sen, A. (1973). Behaviour and the Concept of Preference. *Economica*, *40*(159), 241–259.

Sinervo, B. & Lively, C. M. (1996). The rock–paper–scissors game and the evolution of alternative male strategies. *Nature*, *380*(6571), 240–243.

Skyrms, B. (2000). Game theory, rationality and evolution of the social contract. *Journal of Consciousness Studies*, *7*(1-2), 269–284.

Sugden, R. (1991). Rational Choice: A Survey of Contributions from Economics and Philosophy. *The Economic Journal*, *101*(407), 751–785.

Sugden, R. (2000). Team preferences. *Economics and Philosophy*, *16*, 175–204.

Tadelis, S. (2013). *Game Theory: An Introduction*. Princeton: Princeton University Press.

von Neumann, J., Morgenstern, O., Rubinstein, A., & Kuhn, H. W. (1953). *Theory of Games and Economic Behavior*. Princeton, N.J. ; Woodstock: Princeton University Press.