

# Lecture 06 : Philosophical Issues in Behavioural Science

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

Tuesday, 22nd February 2022

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Expected Utility</b>	<b>2</b>
2.1	Alternative Text . . . . .	2
2.2	Terminology . . . . .	3
<b>3</b>	<b>What Are Preferences?</b>	<b>3</b>
3.1	Required Axioms . . . . .	4
<b>4</b>	<b>Dual Process Theory Opposes Decision Theory?</b>	<b>4</b>
4.1	Background . . . . .	4
<b>5</b>	<b>An Objection to Decision Theory?</b>	<b>4</b>
5.1	How to Write This Week's Assignment . . . . .	4
5.2	The Objection . . . . .	4
5.3	Independence Axiom . . . . .	5
<b>6</b>	<b>Are Objections to Decision Theory also Objections to the Dual Process Theory of Action?</b>	<b>5</b>
<b>7</b>	<b>Question Session 06</b>	<b>6</b>
	<b>Glossary</b>	<b>6</b>

## 1. Introduction

In this part of the course we begin discussion of decision theory (this week) game theory (next week). We will eventually relate these to philosophical issues and psychological discoveries about individual and joint action.

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

- *Instrumental Actions: Goal-Directed and Habitual* in Lecture 01

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

- *Expected Utility* (section §2)
- *What Are Preferences?* (section §3)
- *Dual Process Theory Opposes Decision Theory?* (section §4)

Alternatively, if you have more time but not enough for everything, skip *Dual Process Theory Opposes Decision Theory?* (section §4) and study the other sections.

There is a bit more than usual to cover this week, which will be hard if this is your first encounter with decision theory.

## 2. Expected Utility

The bare minimum you need to know about how actions and rationality are represented in decision theory and in game theory for the purposes of this course.

This section is concerned with understanding the way of representing actions and rationality used in almost any variety of decision theory.

This is not very deep. But you need to understand how the representation of actions is supposed to work in order, later, to understand the theory.

This may well already be familiar ground for you. If so, take a quick look at the slides to check you understand the terminology we will use.

I am mostly following Jeffrey (1983) as this is still the introduction that best combines a deep understanding of the topic with philosophical motivations.

### 2.1. Alternative Text

If you prefer to read a philosopher presenting the core ideas, Bermúdez (2009, chapter 1) is one option. (Bermúdez is summarizing Jeffrey (1983), so read

Jeffrey (1983) if you can.)

## 2.2. Terminology

The recording in this section introduces some terminology that you need to know.

The choice of terms mostly follows Jeffrey (1983), with a few exceptions where his choices are less familiar.

Make sure you understand the terminology and can relate it to the example choice scenario used as an illustration.

Be sure to use the terminology consistently, and with precision, in your writing.

## 3. What Are Preferences?

An informal presentation of Jeffrey (1983, chapter 3) on how decision theory enables us to think of subjective probabilities and preferences as simultaneously derivable from patterns of action.

We have relied on notions of belief and desire in considering both philosophical (in *Philosophical Theories of Action* in Lecture 01) and psychological theories (in *Instrumental Actions: Goal-Directed and Habitual* in Lecture 01) of instrumental action and joint action.

But what anchors our understanding, as researchers, of these notions? While some of us might use these words in everyday life, there is probably enough diversity between individuals with different cognitive styles (e.g.~Perner & Leekam 2008), different upbringings (e.g.~(Morgan et al. 2014)) or different cultural backgrounds (e.g.~(Dixit et al. 2014)) that whatever understandings you and I have in everyday life may not entirely overlap. And invoking a philosophical theory does not seem likely to help given the level of agreement that has been reached in this regard over the last 2000 or so years.

An attractive alternative is suggested by Jeffrey:

This book has ‘a philosophical end: elucidation of the notions of subjective probability and subjective desirability or utility.’ (Jeffrey 1983, xi)

In this section we explore how, following Jeffrey, subjective probabilities and preferences can be identified as constructs of decision theory.

Decision theory therefore promises to be an ideal anchor for a shared understanding of these notions.

Inspired by Jeffery (and Davidson 1990), we might therefore attempt to substitute the informal, poorly understood notions of belief and desire with the theoretical constructs of subjective probability and preference.

### 3.1. Required Axioms

‘A binary relation  $\succsim$  on a set  $A$  is *complete* if  $a \succsim b$  or  $b \succsim a$  for every  $a \in A$  and  $b \in A$ , *reflexive* if  $a \succsim a$  for every  $a \in A$ , and *transitive* if  $a \succsim b$  and  $b \succsim c$ . A preference relation is a complete reflexive transitive binary relation’ (Osborne & Rubinstein 1994, p. 7).

A preference relation is *independent of irrelevant alternatives* exactly if ‘no change in the set of candidates (addition to or subtraction from) [can] change the rankings of the unaffected candidates’ (Dixit et al. 2014, p. 600).

## 4. Dual Process Theory Opposes Decision Theory?

Do any of the findings that support the dual-process theory of instrumental action enable us to construct a good objection to game theory as an elucidation of subjective probabilities and preferences?

### 4.1. Background

The dual-process theory of instrumental action was introduced in *Instrumental Actions: Goal-Directed and Habitual* in Lecture 01.

We considered game theory as an elucidation of subjective probabilities and preferences in *What Are Preferences?* (section §3).

## 5. An Objection to Decision Theory?

This section introduces the Ellsberg Paradox (Ellsberg 1961) and considers how it might be used as an objection to decision theory.

### 5.1. How to Write This Week’s Assignment

Is explained in this section. Step by step.

### 5.2. The Objection

You can hardly pick up a recent work on decision theory without finding an objection to its axioms.

This section introduces an objection linked to the Ellsberg Paradox (Ellsberg 1961; see Hargreaves-Heap & Varoufakis 2004 for an concise and easy to read presentation if you prefer not to watch the recording).

This is just one of many potential objections. I chose it arbitrarily. It gives me an excuse for sharing a fun fact about Ellsberg himself, which illustrates how research in decision making has had life-or-death consequences.

It would be useful to become familiar with other potential objections if you have time. See, for example, Steele & Stefánsson (2020, §2.3) who present the Allais Paradox; or the various objections in Hargreaves-Heap & Varoufakis (2004, Chapter 1); or almost any recent text on decision theory.<sup>[^sug]</sup>

<sup>[^sug]</sup> There are some interesting and influential considerations in Sugden (1991), but this is not the place to start so I recommend considering it only if you already have a good understanding of decision theory and comparatively straightforward objections.

It is perhaps tempting, initially, to think that the objections are simple. They show that decision theory is wrong, misguided or at least too limited to characterise the full richness of human behaviour. But, as we will eventually see, things are much more interesting than that. For it turns out that whether something is an objection depends on what you are using decision theory for.

### 5.3. Independence Axiom

A preference relation is *independent of irrelevant alternatives* exactly if ‘no change in the set of candidates (addition to or subtraction from) [can] change the rankings of the unaffected candidates.’ (Dixit et al. 2014, p. 600)

## 6. Are Objections to Decision Theory also Objections to the Dual Process Theory of Action?

The dual-process theory of instrumental action postulates that, among the processes guiding action selection, there is a goal-directed process. This process is characterised in terms of two representations and their interaction: ‘a representation of the causal relationship between the action and outcome’ and ‘a representation of the current incentive value, or utility, of the outcome’ (Dickinson 2016, p. 177; see *Instrumental Actions: Goal-Directed and Habitual* in Lecture 01). What are these representations? And how does their interaction guide action? As we saw in *What Are Preferences?* (section §3), it is tempting to appeal to decision theory to answer both questions. But, as

we saw in *An Objection to Decision Theory?* (section §5), there are objections to some consturals of decision theory. Are these objections also objections to the combination of decision theory and the dual-process theory of instrumental action?

*The first minute of this section covers the same ground as the start of What Are Preferences? (section §3).*

## 7. Question Session 06

If available (no promises), recordings of the live whole-class lecture will be here, together with slides and references. They are usually available on the day after the session. (You may need to refresh this page to make them appear.)

## Glossary

**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). 2–5

**dual-process theory of instrumental action** instrumental action ‘is controlled by two dissociable processes: a goal-directed and an habitual process’ (Dickinson 2016, p. 177). 4, 5

**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, 4

**instrumental action** An action is *instrumental* if it happens in order to bring about an outcome, as when you press a lever in order to obtain food. (In this case, obtaining food is the outcome, lever pressing is the action, and the action is instrumental because it occurs in order to bring it about that you obtain food.)

You may variations on this definition of *instrumental* in the literature. Dickinson (2016, p. 177) characterises instrumental actions differently: in place of the teleological ‘in order to bring about an outcome’, he

stipulates that an instrumental action is one that is ‘controlled by the contingency between’ the action and an outcome. And de Wit & Dickinson (2009, p. 464) stipulate that ‘instrumental actions are \*learned\*’.  
3, 6

## References

- Bermúdez, J. L. (2009). *Decision Theory and Rationality*. Oxford: Oxford University Press.
- Davidson, D. (1990). The structure and content of truth. *The Journal of Philosophy*, 87(6), 279–328.
- de Wit, S. & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal–human translational models. *Psychological Research PRPF*, 73(4), 463–476.
- Dickinson, A. (2016). Instrumental conditioning revisited: Updating dual-process theory. In J. B. Trobalon & V. D. Chamizo (Eds.), *Associative learning and cognition*, volume 51 (pp. 177–195). Edicions Universitat Barcelona.
- Dixit, A., Skeath, S., & Reiley, D. (2014). *Games of Strategy*. New York: W. W. Norton and Company.
- Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, 75(4), 643–669.
- 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.
- Morgan, G., Meristo, M., Mann, W., Hjelmquist, E., Surian, L., & Siegal, M. (2014). Mental state language and quality of conversational experience in deaf and hearing children. *Cognitive Development*, 29, 41–49.
- Osborne, M. J. & Rubinstein, A. (1994). *A course in game theory*. MIT press.
- Perner, J. & Leekam, S. (2008). The Curious Incident of the Photo that was Accused of Being False: Issues of Domain Specificity in Development, Autism, and Brain Imaging. *Quarterly Journal of Experimental Psychology*, 61(1), 76–89.
- 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.
- 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.
- Steele, K. & Stefánsson, H. O. (2020). Decision Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 ed.). Metaphysics Research Lab, Stanford University.
- Sugden, R. (1991). Rational Choice: A Survey of Contributions from Economics and Philosophy. *The Economic Journal*, 101(407), 751–785.
- 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.