

Lecture 03 : Philosophical Issues in Behavioural Science

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

Wednesday, 22nd October 2025

Contents

1	Introduction	3
2	We Need a Shared Understanding	3
2.1	The Argument	3
2.2	Premise 3: Folk Psychology	4
2.3	Premise 4: Philosophy	5
2.3.1	Limit	6
2.4	Conclusion	6
2.5	Optional Extra: Operationalization	6
3	Expected Utility	7
3.1	Alternative Text	8
3.2	Terminology	8
4	What Are Preferences?	8
4.1	Required Axioms	9
5	Dual Process Theory Opposes Decision Theory?	10
5.1	Background	10
5.2	Argument	10
5.3	Applications of Decision theory	11
5.4	Note on Sources	11
6	An Objection to Decision Theory?	11
6.1	The Objection	11
6.2	How to Object	12
6.3	Independence Axiom	12
6.4	The Paradox of Decision Theory	13

7 Conclusion	13
Glossary	14

1. Introduction

In this lecture we consider decision theory, an attempt to provide a mathematical characterisation of rational behaviour.

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

- *Goal-Directed and Habitual Processes* in Lecture 01

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

- *Expected Utility* (section §3)
- *What Are Preferences?* (section §4)
- *Dual Process Theory Opposes Decision Theory?* (section §5)

Alternatively, if you have more time but not enough for everything, skip *Dual Process Theory Opposes Decision Theory?* (section §5) 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. We Need a Shared Understanding

What anchors our understanding, as researchers, of action, belief, desire and the rest? We examine potential sources for such an understanding—folk psychology and philosophy and find that they cannot anchor a shared understanding. This is where decision theory may be useful: it promises to a formal framework which can elucidate notions of belief and desire, thereby providing a shared understanding.

What anchors our understanding, as researchers, of action, belief, knowledge and the rest?

2.1. The Argument

1. We as researchers need a shared understanding of belief and desire.
2. There are three potential sources of shared understanding: folk psychology, philosophy and decision theory.
3. Folk psychology does not provide a shared understanding.
4. Nor does philosophy.

Therefore:

5. We need decision theory to provide a shared understanding.

To evaluate this argument we need to check premises 3 and 4.

2.2. Premise 3: Folk Psychology

Could our everyday expertise as mindreaders anchor our understanding? Consider Lewis (1972). He postulates that there are a set of platitudes concerning mental states which are common knowledge among us all. He also claims that if we assembled these platitudes, we could use them to define mental state terms like 'intention' and 'knowledge'.

If this were true it would mean that we can rely on our everyday expertise as mindreaders to anchor our understanding, as researchers, of knowledge, intention and the rest. But is it true?

To illustrate how his view works, Lewis imagines that some important platitudes have this form:

'When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and produce so-and-so motor responses.' (Lewis 1972, p. 256)

But what are these platitudes that are supposed to be common knowledge? Heider (1958, p. 12) offered what is probably still, more than half a century later, the most sustained, carefully developed attempt to 'make explicit the system of concepts that underlies interpersonal behavior'.¹ There isn't much in Heider's work that looks useful for defining 'intention' or 'knowledge'.

It is also striking that not very much of Heider's construction could plausibly be regarded as common knowledge among ordinary mindreaders. Heider relies on a mix of informal observation, imagination, guesswork as well as philosophers' ideas (Ryle and Sartre, for example), my guess is that we should regard the principles he identifies not as articulating an understanding that we all share but rather as an imaginative take on possible strategies for everyday mindreading. In fact, Heider's approach is not that different from philosophers like Bratman or Alvarez.

¹ Heider did not share Lewis' assumption about being able to rely on common knowledge of platitudes alone. On Heider's view, 'If people were asked about these conditions they probably would not be able to make a complete list of them. Nevertheless, these assumptions are a necessary part of interpersonal relations; if we probe the events of everyday behavior, they can be brought to the surface and formulated in more precise terms' (Heider 1958, p. 60).

But if Lewis were right about common knowledge of platitudes anchoring mental state terms, either Heider's work should have turned out very differently or else there should be a lot less disagreement among the philosophers. This is why I think Lewis must be wrong.

(We might be able to use theories to specify models that help us characterise the expertise of ordinary mindreaders. But of course this presupposes a shared understanding of belief, desire and the rest in the first place.)

2.3. Premise 4: Philosophy

Could philosophical accounts provide a shared understanding of notions like belief and desire?

One obstacle is widespread apparent disagreement among philosophers.

Take intention, for example. It is not just that philosophers disagree on whether intentions are beliefs about the future (Velleman 1989), or belief-desire pairs (Sinhababu 2013), or something entirely distinct from both beliefs and desires (Bratman 1987). Nor is it just that some think of intentions as essentially components of plans (Bratman 1987 again) whereas others do not connect intentions with plans at all (Searle 1983). Nor is it even that there is much disagreement about how intentions relation to action, to knowledge and to belief. Philosophers even disagree on whether intentions are mental states at all.²

There is similar radical disagreement concerning knowledge, and concerning emotions.

The radical disagreement is an obstacle to using philosophical accounts to anchor understanding just because there is such deep and widespread disagreement among the philosophers.

But there is another, deeper reason for thinking that we cannot use philosophical accounts to anchor our understanding, as researchers, of knowledge, intention and the rest.

Philosophers have different, mostly unarticulated aims. Some philosophers seem to be proposing new ways of thinking in the hope that we adopt them. Others appear to be attempting to make explicit principles that are implicit in a particular tradition of law or in the activities of a particular historical culture. And of course some are trying to make systematic things that seem so obviously true that we can accept them without having any reason to do so (e.g. Lewis 1969).

² 'There is a deep opposition here between accounts that take intention to be a mental state in terms of which we can explain intentional action, and those that do not' (Setiya 2014).

A further problem is that philosophical analysis often appear to rely on researchers' own ideas about folk psychology (returning us to the problems of the previous section). Or so Nagel argues:

'Unless there is a special reason to think that knowledge attributions work quite differently when we are reading philosophy papers—and [there is] evidence against that sort of exceptionalism—we should expect to find that epistemic case intuitions [which are among the things that inform philosophers' views about what knowledge is] are generated by the natural capacity responsible for our everyday attributions of states of knowledge, belief and desire. This capacity has been given various labels, including 'folk psychology', 'mindreading', and 'theory of mind' (Nagel 2012, p. 510).

If Nagel is right, appealing to philosophy to ground a shared understanding also inherits the problems of relying on folk psychology; so it is not even a genuine alternative.

2.3.1. Limit

To be clear, let me distinguish two claims:

1. We could (mis)use philosophical accounts of minds and actions to characterise various models of mind.
2. Philosophical accounts of minds and actions anchor a shared understanding of what knowledge, belief, joy and the rest are.

I am rejecting the second claim only. (The first claim has been very good to me, and I hope to keep misusing philosophical accounts of minds and actions.)

2.4. Conclusion

The argument works: we lack a shared understanding of belief and desire. This is where decision theory may be useful: it promises to a formal framework which can elucidate notions of belief and desire, thereby providing a shared understanding (Jeffrey 1983).

2.5. Optional Extra: Operationalization

A third potential source of shared understanding (in addition to folk psychology and philosophy) is operationalisation. A variety of scientists have

offered operationalisations of abilities to attribute mental states. Perhaps we could use these to anchor a shared understanding?

This is a tempting idea, but it faces two problems. First, we do not really know much about the structure of Theory of Mind, and different researchers use different taxonomies (Happé et al. 2017; Beaudoin et al. 2020).³ Second, while there is some evidence that a wide range of false belief tasks appear to test for a single underlying competence (Flynn 2006, p. 650; Wellman et al. 2001), when we turn to theory of mind tasks more broadly we find that different theory of mind tasks appear to test for different things in the sense that an exploratory factor analysis fails to find that they load on a single factor (Warnell & Redcay 2019).⁴

This means that when faced with the question of what anchors our understanding, as researchers, of action, belief, knowledge and the rest it is not enough simply to point to an operationalisation.

This does not mean, of course, that operationalisations are irrelevant, only that they cannot currently provide a shared understanding of notions like belief and desire.

3. 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.

³ See Beaudoin et al. (2020, p. 15): ‘The lack of theoretical structure and shared taxonomy in ToM definitions and its underlying composition impedes our ability to fully integrate ToM in a coherent and comprehensive framework linking it to various socio-cognitive abilities, a pervasive issue observed across the domain of social cognition.’

⁴ It is important to be clear about why this is a problem. It is not a problem that Theory of Mind may involve a variety of different processes and models, so that no single factor will explain performance across a sufficiently diverse set of tasks. But if you want to say, independently of answering the question about models, that we have a solid operationalization of Theory of Mind, then you need statistics to show that your operationalization has some kind of internal coherence. And that is what appears to be missing.

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

3.1. 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.)

3.2. Terminology

The choice of terms used in this lecture 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.

4. 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 **** ERROR! MISSING xref FOR unit : philosophical_theories_habits ****) and psychological theories (in *Goal-Directed and Habitual Processes* 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. Dixon et al. 2018) 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.⁵

An attractive alternative is suggested by Jeffrey:

⁵ There is a bit more detail on this in some notes for one section of a talk called *The Myth of Mindreading*.

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 Jeffrey (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.

4.1. 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 $b \succsim c$ whenever $a \succsim b$ and $b \succsim c$.

A preference relation is a complete reflexive transitive binary relation’ (Osborne & Rubinstein 1994, p. 7).

The *Continuity Axiom* states that if $c \succ b \succ a$ then there is some probability p such that you are indifferent between (i) b happening with certainty and (ii) a happening with probability p and c happening with probability $(1-p)$.

‘Continuity implies that no outcome A is so bad that you would not be willing to take some gamble that might result in you ending up with that outcome, but might otherwise result in you ending up with an outcome (C) that you find to be a marginal improvement on your status quo (B), provided that the chance of A is small enough.’ (Steele & Stefánsson 2020)

The *Independence Axiom* states that if $b \succ a$ then for any probability p , $\{pA, (1-p)C\} \succsim \{pB, (1-p)C\}$. Put roughly, if you prefer a to b then you should prefer a and c to b and c .

‘Intuitively, this means that preferences between lotteries should be governed only by the features of the lotteries that differ; the commonalities between the lotteries should be effectively ignored.’ (Steele & Stefánsson 2020)

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).

5. 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 decision theory as an elucidation of subjective probabilities and preferences?

5.1. Background

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

We considered decision theory as an elucidation of subjective probabilities and preferences in *What Are Preferences?* (section §4).

5.2. Argument

Consider:

- (1) Decision theory provides an 'elucidation of the notions of subjective probability and subjective desirability or utility' (Jeffrey, 1983, p. xi).
- (2) These notions feature in the goal-directed process, which maximises expected utility.
- (3) Some instrumental actions are dominated by habitual processes.
- (4) Habitual and goal-directed processes can pull in opposing directions.

therefore:

- (5) Some actions do not maximise expected utility.

therefore:

- (6) Premise (1) is false (decision theory cannot elucidate subjective probability and subjective desirability).

This argument has the form of a reductio.

The main conclusion of the argument, (6), is significant if true: it suggests that we cannot use decision theory as an anchor for thinking about notions of belief and desire. But perhaps there is a way to avoid this conclusion?

Are the premises of the argument true? Or is there some way to use decision theory as an anchor for thinking about notions of belief and desire despite the inconsistency of the above claims?

5.3. Applications of Decision theory

The above argument is an objection to a particular application of decision theory. It is not ‘an objection to decision theory’ because there is no such thing. Decision theory is not the kind of thing that you can object to ...

Decision theory is a model. Like any model, it can be given different applications. There are no objection to decision theory as such, which is simply a model. Instead each objection is an objection to one or more applications of decision theory.

The above objection is an objection to applying decision theory to elucidate notions of subjective probability and preference. It is not an objection to applying decision theory to predict actions, nor to applying decision theory to characterise ideally rational actions.

5.4. Note on Sources

One possible response to the above argument discussed in the lecture involves a distinction between between computational description and implementation details. This is a rough-and-ready approximation to a famous three-fold distinction from Marr (1982); in terms of that theory my ‘implementation details’ are what Marr calls representations and algorithms.

6. 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.

This is an optional section that was not covered in all versions of the lecture this year.

6.1. The Objection

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

This section introduces on 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.⁶

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.

6.2. How to Object

0. State the construal of decision theory you are considering.
1. State the finding.
 - (In this case, the finding is Ellsberg's discovery of cases where people prefer A over B but also prefer B or C over A or C.⁷)
2. State the axiom it contradicts.
3. Explain how the finding contradicts the axiom.
4. (If possible, explain why it is significant.)
5. Consider responses.

6.3. Independence Axiom

The *Independence Axiom* states that if $b \succsim a$ then for any probability p , $\{pA, (1-p)C\} \succsim \{pB, (1-p)C\}$. Put roughly, if you prefer a to b then you should prefer a and c to b and c.

'Intuitively, this means that preferences between lotteries should be governed only by the features of the lotteries that differ; the commonalities between the lotteries should be effectively ignored.' (Steele & Stefánsson 2020)

⁶ 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.

⁷ You can also mention Jia et al. (2020)'s findings if you are being especially thorough.

6.4. The Paradox of Decision Theory

On the one hand, it has become a commonplace that there are plenty of objections to the idea that decision theory characterises how people choose.

On the other hand, there is a growing range of cases in which decision theory (or something based on it, like game theory) has been fruitfully applied. Motor control is a prominent example (see Trommershäuser et al. 2009; Wolpert & Landy 2012).

If the objections are as decisive as usually assumed, why have applications of decision theory proved so fruitful?

Perhaps the answer is that decision theory is a model. Like any model, it can be given different construals. The objections are not objections to decision theory as such, which is simply a model. Instead each objection is an objection to one or more construals of decision theory.

If this is right, it will be important to be clear about which construals your objections concern.

7. Conclusion

After this lecture you should understand what decision theory is, why we need something to anchor a shared understanding among us, as researchers, of the notions of belief and desire, why it is at least theoretically coherent to construe decision theory as providing this, and why construing decision theory in this way is difficult or impossible to combine with accepting the dual-process theory of instrumental action.

The overall aim of the course: to discover why people act, individually and jointly.

To have any chance of achieving this, we need a synthesis of:

- the kind of theoretical framework provided by philosophical thinking;
- a body of evidence provided by experimental psychology; and
- a formal model.

At this point, we have considered all three items.

This lecture was about the formal models. The best studied, most influential of these is decision theory.

Why do we need decision theory, and how does it fit with the philosophical and psychological theories considered so far?

One possibility is that decision theory provides an elucidation of the notions of belief and desire that we need to characterise goal-directed processes (Jeffrey 1983); see *What Are Preferences?* (section §4).

But, as we saw in *Dual Process Theory Opposes Decision Theory?* (section §5), it is not straightforward to combine this idea with the dual-process theory of instrumental action.

Glossary

anchor A theory, fact or other thing that is used by a group of researchers to ensure that they have a shared understanding of a phenomenon. An anchor is needed when it is unclear whether different researchers are offering incompatible claims about a single phenomenon or compatible claims about distinct phenomena. For example, we might take decision theory to anchor a shared understanding of belief and desire. 8, 10, 11

computational description A computational description of a system or ability specifies what the thing is for and how it achieves this. Marr (1982) distinguishes the computational description of a system from representations and algorithms and its hardware implementation. 11, 15

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). 3, 8–11, 13, 14

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). (See instrumental action.) 10, 13, 14

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). 13

goal-directed process A process which involves ‘a representation of the causal relationship between the action and outcome and a representation of the current incentive value, or utility, of the outcome’ and

which influences an action ‘in a way that rationalizes the action as instrumental for attaining the goal’ (Dickinson 2016, p. 177). 14

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 encounter variations on this definition of *instrumental* in the literature. For instance, 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*’. 8, 14

model A model is a way some part or aspect of the world could be. 11, 13

representations and algorithms To specify the representations and algorithms involved in a system is to specify how the inputs and outputs are represented and how the transformation from input to output is accomplished. Marr (1982) distinguishes the representations and algorithms from the computational description of a system and its hardware implementation. 11, 14

References

- Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Frontiers in Psychology, 10*.
- Bermúdez, J. L. (2009). *Decision Theory and Rationality*. Oxford: Oxford University Press.
- Bratman, M. E. (1987). *Intentions, Plans, and Practical Reasoning*. Cambridge, MA: Harvard 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.
- Dixson, H. G. W., Komugabe-Dixson, A. F., Dixson, B. J., & Low, J. (2018). Scaling Theory of Mind in a Small-Scale Society: A Case Study From Vanuatu. *Child Development*, *89*(6), 2157–2175.
- Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, *75*(4), 643–669.
- Flynn, E. (2006). A microgenetic investigation of stability and continuity in theory of mind development. *British Journal of Developmental Psychology*, *24*(3), 631–654.
- Happé, F., Cook, J. L., & Bird, G. (2017). The Structure of Social Cognition: In(ter)dependence of Sociocognitive Processes. *Annual Review of Psychology*, *68*(1), 243–267.
- Hargreaves-Heap, S. & Varoufakis, Y. (2004). *Game Theory: a Critical Text*. London: Routledge.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. Hillsdale, N.J: Lawrence Erlbaum Associates.
- Jeffrey, R. C. (1983). *The Logic of Decision, second edition*. Chicago: University of Chicago Press.
- Jia, R., Furlong, E., Gao, S., Santos, L. R., & Levy, I. (2020). Learning about the Ellsberg Paradox reduces, but does not abolish, ambiguity aversion. *PLOS ONE*, *15*(3), e0228782.
- Lewis, D. K. (1969). *Convention : a philosophical study*. Cambridge, MA.: Harvard University Press.
- Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, *50*(3), 249–258.
- Marr, D. (1982). *Vision : a computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
- 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.

- Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research*, 85(3), 495–527.
- 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.
- Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Sen, A. (1973). Behaviour and the Concept of Preference. *Economica*, 40(159), 241–259.
- Setiya, K. (2014). Intention. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.). Metaphysics Research Lab, Stanford University.
- Sinhababu, N. (2013). The Desire-Belief Account of Intention Explains Everything. *Noûs*, 47(4), 680–696.
- 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.
- Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2009). Chapter 8 - The Expected Utility of Movement. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics* (pp. 95–111). London: Academic Press.

Velleman, D. (1989). *Practical Reflection*. 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.

Warnell, K. R. & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, 191, 103997.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, 72(3), 655–684.

Wolpert, D. M. & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, 22(6), 996–1003.