

Lecture 01 : Philosophical Issues in Behavioural Science

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

Monday, 30th September 2024

Contents

1	Introduction: Why Investigate Philosophical Issues in Behavioural Science?	2
1.1	Overview	2
1.2	The Simple Picture of Why People Act	2
1.3	Structure: Two Questions	2
1.4	Integration Question	3
2	Practical Things	3
3	Philosophical Theories of Action	3
3.1	The Problem of Action	3
3.2	Why It Matters	4
3.3	A Standard Solution	5
3.4	Appendix: Davidson on Agency	5
4	Instrumental Action	7
4.1	Terminology	7
4.2	Main Question	7
4.3	A Standard Answer	7
5	Goal-Directed and Habitual Processes	8
5.1	A Clue from Animal Learning	9
5.2	What Are Habitual Processes?	10
5.3	Habitual Processes and Instrumental Action	10
5.4	So What?	11
6	Conclusion	11
	Glossary	12

1. Introduction: Why Investigate Philosophical Issues in Behavioural Science?

Answering broadly philosophical questions about action and about joint action requires reflection on discoveries from the behavioural sciences.

1.1. Overview

To discover why people act, individually and jointly. This is the overall aim of the course.

Our premise is that discovering why people act requires multiple methods: philosophical, psychological and formal.

An obstacle is that these three methods—philosophical, psychological and formal—are associated with different theories and the theories seem to be inconsistent with each other.

1.2. The Simple Picture of Why People Act

Start with a simple picture that is widely assumed in philosophical and economic theories:

When you act, there are reasons why you act; you know the reasons; you act because you know the reasons; and the reasons justify your action.

Is this true? How can we turn it into a theory? Are there alternative pictures?

Much of what we will study are attempts to answer these questions.

1.3. Structure: Two Questions

Here's the structure of the course. Two questions. One about individual action, the other about collaborative action:

1. Which events in your life are your actions?
2. What distinguishes doing something jointly with another person from acting in parallel with them but merely side by side?

We start with the first question then move on to the second.

We will consider each question from three perspectives: philosophical, psychological and formal.

Although the perspectives are distinct, they are not independent. As things stand, we cannot adequately answer broadly philosophical questions about action, nor about joint action, without reflection on discoveries from the behavioural sciences.

1.4. Integration Question

The course centers on a set of *Integration Questions*:

where there are philosophical, psychological and formal theories which appear to target a single set of phenomena while saying incompatible things about it, we face two questions: * are they actually inconsistent? * if so: how, if at all, should either or both theories be refined?

As we will see, apparent conflicts between philosophical, psychological and formal theories arise at the most fundamental level.

2. Practical Things

Quick overview of seminars and assessment methods. You can ask questions here if anything is unclear.

I put the notes for this part at the top of the page: Practical Information

3. Philosophical Theories of Action

Much philosophy of action starts with The Problem of Action: What distinguishes your actions from things that merely happen to you (Davidson, 1971)? According to a standard, widely-accepted solution, actions are those events which stand in an appropriate causal relation to an intention. This is an instance of the Causal Theory of Action, according to which an event is action 'just in case it has a certain sort of psychological cause' (Bach, 1978, p. 361).

The recording is for an older version of this unit with different content.

3.1. The Problem of Action

Much philosophy of action hinges on the question, What distinguishes your actions from things that merely happen to you (Davidson 1971)?¹

¹ See Shepherd (2021, p. 1): 'The history of philosophical reflection on action gives the distinction between activity and passivity different names, and attempts to explain the

You trip and fall down a flight of stairs. Falling is something that happens to you, not an action of yours. But watching the sympathetic attention you gain, Buster expertly throws himself down the stairs. Although it looks like another accident, this event is an action.

As Frankfurt (1978, p. 157) put it:

‘The² problem of action is to explicate the contrast between what an agent does and what merely happens to him.’

But is this really a problem? It may be tempting, initially, to suppose that we can answer this question by invoking kinematic features. Perhaps—so the idea—actions are those events which involve some or other patterns in the joint displacements and bodily configurations? Alternatively, it might be tempting to think that we can answer the question by appeal to coordination. Perhaps—so the thought—actions are those events which involve a particular coordination of body parts? If either possibility obtained, the ‘problem of action’ would not be a problem at all. But reflection on the variety of things that count as actions indicates that neither of these initially tempting possibilities is at all likely to obtain. Or so I argue in *Recap: Action* from the lectures on Mind and Reality.

The absence of straightforward answers to the question about what distinguishes actions from things that merely happen to you indicates that it is a genuine problem.

3.2. Why It Matters

Our overall concern on this course is to understand why people act, individually and jointly.

To see why the The Problem of Action matters, suppose we just replace ‘act’ with ‘move’.

So many things move for so many different reasons—rocks, people, plants, continental plates and bacteria—that it makes no sense to look for a general theory about why things move.

If we are to have a coherent research project, we need a principled way of limiting our enquiry to actions as opposed to movements more generally.

Bacteria turn out to behave in suprisingly sophisticated way, as do plants and, of course, all kinds of machines.

distinction in different ways. But philosophers circle the distinction repeatedly [...]’ Shepherd goes on to mention several famous historical sources for The Problem of Action.

² I dislike this way of stating things. Good philosophers come up with lots of questions. There is insufficient reason to single one of them out as *the* problem.

The Problem of Action matters because a solution to it will be a principled way of delimiting the things we are asking about so that we are not thinking above movement generally.

3.3. A Standard Solution

According to a standard, widely-accepted view, actions are those events which stand in an appropriate causal relation to an intention. What distinguishes your falling from Buster's is that his, but not yours, was appropriately related to an intention.

This is an instance of the Causal Theory of Action. According to this view, an event is action 'just in case it has a certain sort of psychological cause' (Bach 1978, p. 361). Proponents of this view may disagree about which states cause actions (Bach is an example of this), or about how to characterise the causal relation (for example, Frankfurt (1978) is concerned, in part, with whether the causes are antecedent to the action or provide ongoing guidance). But they agree that the relation between actions and their psychological causes is what distinguishes your actions from things that merely happen to you.

3.4. Appendix: Davidson on Agency

This is an optional extra section. It is not part of the lecture.

How does Davidson arrive at the view that actions are those events which stand in an appropriate causal relation to an intention?³

As background, Davidson notes that the same action can be described in multiple ways. You move your finger, flicking a switch which causes the lights to come on and alerts a prowler (Davidson 1971, p. 53). We have four ways of describing one and the same action: as moving your finger, as flicking a switch, and so on.

Davidson further notes that actions can typically be described both in ways that relate to what you intended (turning the lights on, say) and in ways which do not relate to your intentions (alerting a prowler, perhaps).

This background allows Davidson to distinguish three situations involving someone spilling coffee:

³ I've heard people who should know say that Davidson does not explicitly commit to this view. But Davidson writes, 'we have discovered no analysis of this relation that does not appeal to the concept of intention' (Davidson 1971, p. 61). And nowhere does he explicitly reject the view that actions are those events which stand in an appropriate causal relation to an intention.

'If [...] I intentionally spill the contents of my cup, mistakenly thinking it is tea when it is coffee, then spilling the coffee is something I do, it is an action of mine, though I do not do it intentionally. On the other hand, if I spill the coffee because you jiggle my hand, I cannot be called the agent. Yet while I may hasten to add my excuse, it is not incorrect, even in this case, to say I spilled the coffee. Thus we must distinguish three situations in which it is correct to say I spilled the coffee: in the first, I do it intentionally; in the second I do not do it intentionally but it is my action (I thought it was tea); in the third it is not my action at all (you jiggle my hand).' (Davidson 1971, p. 45)

In short my spilling the coffee can be caused in three ways:

1. by an intention of mine to spill the coffee;
2. by an intention of mine to spill the tea (where I mistakenly take the coffee to be tea and do not intend to spill coffee);
or
3. by you jiggling my hand (where no intention of mine is directly involved at all).

My spilling the coffee is an action of mine in (1) and (2), but not in (3).

Reflection on (1) and (2) rules out the view that my spilling the coffee is an action of mine only if I intend to spill the coffee.

The contrast between (2) and (3) is what leads Davidson to his view about agency:

'What is the difference [between (2) and (3)]? The difference seems to lie in the fact that in one case, but not in the other, I am intentionally doing something. My spilling the contents of my cup was intentional; as it happens, this very same act can be redescribed as my spilling the coffee. Of course, thus redescribed the action is no longer intentional; but this fact is apparently irrelevant to the question of agency.

'And so I think we have one correct answer to our problem: a man is the agent of an act if what he does can be described under an aspect that makes it intentional.' (Davidson 1971, p. 46)

Suppose we assume, further, that an act can be described under an aspect that makes it intentional only if it stands in an appropriate causal relation to an intention of the agent's.⁴ Then the Standard Solution mentioned above

⁴ Is this assumption true? Bratman allows that actions can be intentional 'even though

follows:

Your actions are those events which stand in an appropriate causal relation to an intention of yours.

4. Instrumental Action

An instrumental action is an action that happens in order to bring about an outcome. When you press a lever in order to retrieve a snack, or when you board a bus in order to travel home, you are performing an instrumental action. What grounds the relation between an instrumental action and the outcome it occurs in order to bring about?

4.1. Terminology

An instrumental action is an action that happens in order to bring about an outcome. We will say that the outcome is a goal of the action,⁵ and that the action is directed to the outcome.

4.2. Main Question

What is the relation between an instrumental action and the outcome or outcomes to which it is directed?

4.3. A Standard Answer

One standard answer to this question involves intention. An intention specifies an outcome, coordinates your actions, and coordinates your actions in a way that would normally increase the probability of the specified outcome occurring. So if an intention causes you to act, it follows that your action happens in order to bring about the outcome intended. And this implies that your action is instrumental.

[the agent] has no distinctive attitude of intending' (Bratman 1987, p. 132), and even though the agent lacks the capacity to form intentions altogether (Bratman 2000, p. 51). This view follows from two claims: first, intentions are distinct from any combination of beliefs and desires; and second, beliefs and desires alone may, in certain cases, determine what an agent intentionally does.

⁵ Be careful not to confuse a goal with a goal-state, which is an intention or other state of an agent linking an action to a particular goal to which it is directed. (Some authors use the term 'goal' for goal-states rather than outcomes.) A goal is a possible or actual outcome (such as filling a glass with prosecco). A goal-state is a psychological attribute of an agent (such as an intention to fill a glass with prosecco).

What is an intention? Although there is much debate about this (Setiya 2014), for our purposes only a widely agreed characteristic is necessary. Intentions are the upshot of beliefs and desires (or are identical to one or both of these). To illustrate:

desire: I fill Zak's glass.

belief: If I pour, I will fill Zak's glass.

intention: I pour to fill Zak's glass.

This simplistic example captures a key idea. Behind an intention lie two things. There is a desire to bring an outcome about, and there is a belief about which action will bring the action about.⁶

If you would like more background on action and intention, see Lecture 10 of *Mind and Reality*.

Our Main Question is about the relation between an instrumental action and the outcome or outcomes to which it is directed. According to the Standard Answer, the relation involves belief, desire and intention:

Background Assumption: Instrumental actions are caused by intentions to bring outcomes about, which are the upshot of desires to bring outcomes about and beliefs that certain actions will bring them about.

Standard Answer: The outcome (or outcomes) to which an instrumental action is directed is that outcome (or outcomes) specified by the intention (or intentions) which caused it.

Does the Standard Answer involving intention provide a full answer to that question? Or are there things other than intentions which might link an instrumental action to an outcome? The next section provides a reason for thinking that there are.

5. Goal-Directed and Habitual Processes

This section introduces a key distinction between goal-directed and habitual processes.

In *Instrumental Action* (section §4), we asked about the relation between an instrumental action and the outcome or outcomes to which it is directed. As we saw, the Standard Answer given by philosophers is that intention grounds this relation.

⁶ We will see the same structure when we come to decision theory (in *Expected Utility* in Lecture 03). Preferences correspond to desires and expected utilities to beliefs.

But are there maybe things other than intentions which might link an instrumental action to an outcome?

5.1. A Clue from Animal Learning

According to Dickinson (2016, p. 177):

‘instrumental behavior is controlled by two dissociable processes: a goal-directed and an habitual process’

He goes on to specify what the ‘goal-directed process’ involves:

‘an action is goal-directed if it is mediated by the interaction of a representation of the causal relationship between the action and outcome and a representation of the current incentive value, or utility, of the outcome in a way that rationalizes the action as instrumental for attaining the goal’ Dickinson (2016, p. 177).

Dickinson’s ‘goal-directed process’ corresponds to the belief–desire model we just considered. The ‘representation of the causal relationship between the action and outcome’ could be a belief about which action will bring an outcome about (e.g. the belief that if I pour, I will fill Zak’s glass). And the ‘representation of the current incentive value, or utility, of the outcome’ could be a desire.

philosophy	animal learning	decision theory
belief	representation of the causal relationship between the action and outcome	subjective probability
desire	representation of the current incentive value, or utility, of the outcome	preference

Table: rough correspondence between terms used for modelling action across three disciplines.

So when Dickinson says that instrumental actions are ‘controlled by two dissociable processes’, he is implying that the Standard Answer about belief, desire and intention cannot fully explain the relation between an instrumental action and the outcome or outcomes to which it is directed. If he is right, we also have to consider something he calls ‘an habitual process’.

5.2. What Are Habitual Processes?

Habitual processes involve connections between stimuli and actions. For example, the presence of an empty glass (a stimulus) may be connected to the action of pouring. These connections are characterised by two features:

1. When the action is performed in the presence of the stimulus, the connection between action and stimulus is strengthened (or 'reinforced') if the action is rewarded.
2. If the connection is strong enough, the presence of the stimulus will cause the action to occur.

This is another way of stating *Thorndyke's Law of Effect*:

'The presentation of an effective [=rewarding] outcome following an action [...] reinforces a connection between the stimuli present when the action is performed and the action itself so that subsequent presentations of these stimuli elicit the [...] action as a response' (Dickinson 1994, p.48).

How do habitual processes differ from those involving belief, desire and intention? Two differences are important for our purposes:

1. The effects of habitual processes do not depend on what you currently desire. This is because the strength of the stimulus–action connection depends only on what was rewarding for you in the past, not what is rewarding for you now.
2. The effects of habitual processes do not depend on what you currently believe about which outcome the action will have. This is because the strength of the stimulus–action connection depends only on what outcomes the action had in the past, not on which outcomes it will have now.

Because habitual processes have these features, we can be sure that they are genuinely distinct from processes involving belief, desire and intention.

5.3. Habitual Processes and Instrumental Action

Our Main Question is, What is the relation between an instrumental action and the outcome or outcomes to which it is directed? This question can be answered by invoking habitual processes. For if an action is due to an habitual process, then there is a stimulus–action connection which caused it. This stimulus–action connection must have been strengthened in the past because, often enough, some (one or more) rewarding outcomes occurred

when the action was performed in the presence of the stimulus. But since habitual processes exist to enable the agent repeatedly bring about such rewarding outcomes, it follows that the action occurs now in order to bring about these (one or more) rewarding outcomes. That is, the action is directed to the outcome; it is an instrumental action.

The Standard Answer therefore fails to provide a full answer to the Main Question about instrumental action. To fully answer it we need not only belief, desire and intention but, minimally, also the kind of stimulus–action connections involved in habitual processes.

5.4. So What?

After this section, you should understand what an instrumental action is, you should understand the Main Question, and you should understand how habitual processes and goal-directed processes differ.

The next step is to investigate possible consequences for philosophical theories of action.

6. Conclusion

In this lecture we have begun to think about instrumental action from the point of view of theories of animal learning, distinguishing habitual from goal-directed processes. And we have considered action from the point of view of philosophy of action, focussing on The Problem of Action and the notion of intention.

The challenge for the whole course is to discover why people act, individually and jointly.

In this lecture we encountered two questions about action:

Question 1: What is the relation between an instrumental action and the outcome or outcomes to which it is directed? (see *Goal-Directed and Habitual Processes* (section §5))

Question 2: What distinguishes your actions from things that merely happen to you? (The Problem of Action, see *Philosophical Theories of Action* (section §3))

Philosophers standardly answer both questions by invoking intention. This supports the Simple Picture of why people act. But is it the whole story?

On Question 1, we can coherently answer the first question by appeal to habitual processes without invoking intention at all (see *Goal-Directed and*

Habitual Processes (section §5)). This suggests that the standard philosophical answer is not the whole story.

Our next step will be to examine whether the existence of habitual processes creates a problem for philosophical answers to the second question.

Glossary

Causal Theory of Action According to this view, an event is action ‘just in case it has a certain sort of psychological cause’ (Bach 1978, p. 361). 5

directed For an action to be *directed* to an outcome is for the action to happen in order to bring that outcome about. 7, 10, 11

goal A *goal* of an action is an outcome to which it is directed. 7

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). 11

goal-state an intention or other state of an agent which links an action of hers to a particular goal to which it is directed. 7

habitual process A process underpinning some instrumental actions which obeys *Thorndyke’s Law of Effect*: ‘The presentation of an effective [=rewarding] outcome following an action [...] reinforces a connection between the stimuli present when the action is performed and the action itself so that subsequent presentations of these stimuli elicit the [...] action as a response’ (Dickinson 1994, p.48). (Interesting complication which you can safely ignore: there is probably much more to say about under what conditions the stimulus–action connection is strengthened; e.g. Thrailkill et al. 2018.) 10–12

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*’. 7–11

outcome An outcome of an action is a possible or actual state of affairs. 7–11

problem a question that is difficult to answer. 4

stimulus A *stimulus* is just a situation or event. Typically, ‘stimulus’ is used to label things which do, or might, prompt actions such as the presence of a lever or the flashing of a light. 10

The Problem of Action What distinguishes your actions from things that merely happen to you? (According to Frankfurt (1978, p. 157), ‘The problem of action is to explicate the contrast between what an agent does and what merely happens to him.’) 4, 5, 11

References

- Bach, K. (1978). A representational theory of action. *Philosophical Studies*, 34(4), 361–379.
- Bratman, M. E. (1987). *Intentions, Plans, and Practical Reasoning*. Cambridge, MA: Harvard University Press.
- Bratman, M. E. (2000). Valuing and the will. *Noûs*, 34(supplement 14), 249–265. Reprinted in Bratman, M. (2007) *Structures of Agency*. Oxford: Oxford University Press (pp. 47–67).
- Davidson, D. (1971). Agency. In R. Binkley, R. Bronaugh, & A. Marras (Eds.), *Agent, Action, and Reason*, (pp. 3–25). Toronto: University of Toronto Press. Reprinted in Davidson, D. (1980) *Essays on Actions and Events*. Oxford: Oxford University Press.
- 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. (1994). Instrumental conditioning. In N. Mackintosh (Ed.), *Animal Learning and Cognition*. London: Academic Press.
- 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.

- Frankfurt, H. (1978). The problem of action. *American Philosophical Quarterly*, 15(2), 157–162.
- Setiya, K. (2014). Intention. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.). Metaphysics Research Lab, Stanford University.
- Shepherd, J. (2021). *The Shape of Agency: Control, Action, Skill, Knowledge*. Oxford: Oxford University Press.
- Thrailkill, E. A., Trask, S., Vidal, P., Alcalá, J. A., & Bouton, M. E. (2018). Stimulus control of actions and habits: A role for reinforcer predictability and attention in the development of habitual behavior. *Journal of Experimental Psychology: Animal Learning and Cognition*, 44, 370–384.