Introduction

The aim of this book is to solve a problem facing contemporary metaphysical theories of action and agency.

The central task of a theory of action is to explain the difference between action and mere behaviour. When I raise my arm in order to hail a passing cab, the raising of my arm isn’t something that merely happens to me; rather, I exercise control over my behaviour, and thereby exercise my agency. By contrast, when my arm rises as the result of an involuntary spasm, the rising of my arm is something that merely happens to me; I don’t exercise control over my behaviour, and so I don’t exercise my agency. The central task of a theory of action is to explain what the difference between these different sorts of behaviour consists of. For instance, according to the Causal Theory of Action defended by Donald Davidson and others, the difference between actions and mere behaviours is a difference in their causal histories. When I raise my arm in order to hail a passing cab, the rising of my arm is caused (in the right way) by certain mental states of mine, namely, intentions, beliefs, and desires which rationalize my behaviour. Because it is caused in this way, the rising of my arm counts not merely as an arm-raising, but as an arm-raising. By contrast, when my arm rises as the result of a spasm, it lacks the sort of causal history that an arm-raising has, and so it is merely an arm-raising. The hope, of course, is that a similar story can be told for all sorts of actions, not just arm-raising.

Now, it’s an important fact about our agency that we can exercise it both by doing things and by not doing things. If I intentionally omit to pick up a friend at the airport, or refrain from having a second helping of dessert, I thereby exercise my agency, just as when I raise my arm in order to hail a passing cab. When I intentionally omit to do something, or refrain from doing it, I thereby exercise control over my own behaviour, and so it seems that I’ve acted, in omitting or refraining. By contrast, if I simply forget to pick up my friend at the airport, or fail to have a second.
2  Introduction

helping of dessert simply because it doesn’t occur to me to take seconds, I haven’t exercised my agency; I have simply not done something.

A satisfactory theory of action and agency must accommodate ‘negative actions’ – i.e. exercises of agency which seem to consist primarily in an agent not doing a certain thing. Such behaviours are precisely the sort of thing that a theory of action is supposed to be a theory of.

However, negative actions present a metaphysical problem for the dominant theories. These theories – including the Causal Theory – are ‘event-based’: they are built on the assumption that actions are events, i.e. occurrences or happenings. With this assumption in place, the central task of a metaphysical theory of action becomes the task of distinguishing actions from other events, i.e. explaining what distinguishes actions as a subset of the larger category of events. The problem is that, according to many philosophers of action, negative actions (with perhaps a few exceptions) are not events. For, it seems, if doing something is a matter of the occurrence of an event, then not doing that thing is a matter of the absence of an event: if I intentionally omit to raise my arm, what’s happened isn’t that an event of a special kind, an omission to raise my arm, has occurred; rather, what’s happened is simply that no event of the kind raising of my arm has occurred. The problem is clear. If all actions are events, but many (and perhaps all) negative actions aren’t, then so-called negative actions aren’t really actions at all, but absences of action. The dominant, event-based theories of action seem incapable of accommodating an important class of actions, and hence incapable of accommodating an important aspect of our agency.

Call this ‘the problem of negative action’. It can be presented as an inconsistent triad. Let ‘NEG’ be a schematic verb for negative behaviours, e.g. omitting and refraining, which can be combined with another verb, ‘φ’, to generate a negative verb phrase, ‘NEG-φ’, whose instances include ‘omit to φ’, ‘refrain from φ-ing’, and the like. Then, each of the following propositions is deeply intuitive, or widely held among philosophers of action for theoretical reasons, or both, but they can’t all be true.

(PNA 1) Negative actions – in the sense of things done – are genuine actions; necessarily, if NEG-φ is a negative action for x at t, then if x NEG-φ-s at t, then x acts by NEG-φ-ing at t.

(PNA 2) Necessarily, if x acts by φ-ing at t, then there exists an event that is x’s token φ-ing at t.

(PNA 3) Possibly, x NEG-φ-s at t, and there exists no event that is x’s token NEG-φ-ing at t.

One or more of these propositions must be rejected.
To reject (PNA 1) is to deny the appearances and insist that so-called negative actions aren’t really actions at all. If we go this way, then we needn’t revise or reject the dominant event-based theories in light of the alleged counterexamples of intentional omissions, refrainments, and the like. If these behaviours aren’t actions, then they aren’t the sort of thing that a theory of action is meant to be a theory of, and so they can’t be counterexamples to event-based theories.

To reject (PNA 2), by contrast, is to take on the burden of revising the dominant event-based theories so as to free them from the assumption that actions are events – e.g. in the case of the Causal Theory, providing an account of how there can still be a causal process of the right kind when there is no such behaviour as my omission to raise my arm, to which mental states of mine can stand in the relation of singular causation – or rejecting those theories altogether.

The solution I prefer is to reject (PNA 3): negative actions are events, not absences. If we go this way, then we get the benefit of the first solution: we needn’t revise or reject the dominant event-based theories in order to accommodate negative actions; since negative actions are events, rather than absences thereof, then they pose no metaphysical problem for event-based theories. And we do so without having to deny what seems prima facie obvious, that intentional omissions, refrainments, and the like are exercises of agency, and that a theory of action should be able to explain the difference between intentionally omitting to do something, or refraining from doing it, on one hand, and merely not doing it, on the other.

Despite these attractions, my preferred solution has few defenders. The bulk of the book will be dedicated to addressing the motivations behind (PNA 3), showing that they fail, and developing an account of negative actions as events which will answer the standard worries one finds in the literature.

In defending my solution, I appeal in part to ideas and arguments from the philosophy of language. The major motivation for (PNA 2) comes from the ‘Neo-Davidsonian’ approach to the analysis of action sentences. According to this approach, ordinary action sentences such as ‘I raised my arm’ are existential quantifications over events: the logical form of ‘x φ-s’ is ‘There exists an event, e, that is x’s φ-ing’. Thus, these sentences are true just in case those events exist. The major motivation for (PNA 3) is that the same is not true for negative action sentences. According to a widely-assumed view which I call ‘Deflationism’, negative action sentences are simply the negations of corresponding positive action sentences, and hence they express negative existentials: the logical form of ‘x NEG-φ-s’ is ‘There
4 Introduction

exists no event, e, which is x’s φ-ing’. There’s no requirement that there be any such event as x’s NEG-φ-ing in order for this sentence to be true; what’s required is the absence of a certain kind of event, not the presence of one. Hence (PNA 3) and the idea that negative actions are not events but absences thereof.

I argue that De falsity is false. The best account of the logical form of negative action sentences is a Neo-Davidsonian one, according to which they report the occurrence of negative actions quas events. Thus, the motivation for thinking that ordinary actions are events extends to negative ones as well.

This style of argument relies on a Quinean criterion of ontological commitment. According to this criterion, you’re ontologically committed to the entities which are quantified over in the sentences you accept: if you accept a sentence ‘S’ (equivalently: if you think that S is the case) and ‘S’ has the form ‘∃xFx’ (and so, ‘S’ is true just in case there are Fs), then you’re ontologically committed to Fs. Thus, if negative action sentences quantify over negative actions qua events, and you think those sentences are sometimes true (that is, if you think that people sometimes omit to do things, refrain from doing things, etc.), then you ought to think that negative actions are events, not absences.

This criterion, and this style of argument, are unpopular these days. Many metaphysicians reject the Quinean criterion, and even deny that we can appeal to considerations of language and logical form at all when engaging in properly metaphysical argument. I defend the Quinean criterion in more detail in Section 1.3.2, but some remarks about it, and the style of argument I’ll be giving in later chapters, are in order here.

It’s often thought that if we appeal to considerations of language and logical form when arguing for a metaphysical conclusion then we fail to do justice to the metaphysician’s conception of herself as inquiring into the nature of reality, rather than into how we happen to talk about it. Worse, it’s often thought that to appeal to considerations of this sort is to attempt to ‘read off’ facts about how things are from facts about how we talk, or to engage in a shaky inference from ‘We speak as if p’ to p. But we needn’t reduce metaphysics to ‘mere’ linguistics or take how we happen to talk as an infallible guide to how things are in order to make the kind of argument I’ll be making in this book.

Consider the following argument form:

(α) ‘S’ is true $\equiv S$.
(β) ‘S’ is true $\equiv p$.
(γ) :. S $\equiv p$.

(α) is simply the T-schema, according to which you can give the truth-conditions for a sentence by using that sentence – e.g. ‘Snow is white’ is...
true just in case snow is white. (β) is a schema for giving the truth-conditions for a sentence without using that very same sentence. ‘p’ names the propositional content of ‘S’ and needn’t be structurally equivalent to ‘S’ – e.g. using the resources of first-order formal languages we might say that ‘Snow is white’ is true just in case White (snow). If both (α) and (β) are true, then (γ) follows by elementary reasoning: the state of affairs reported by ‘S’ obtains just in case the state of affairs reported by ‘p’ obtains.

(α) and (β) are framed in meta-linguistic terms, and so there’s a clear sense in which they are claims about language. By contrast, unless ‘S’ itself is a meta-linguistic sentence, (γ) isn’t about language in any obvious sense; it reports that one state of affairs obtains just in case another state of affairs obtains. Nonetheless, since the argument is valid, there’s no reason to think that (γ) can’t follow from (α) and (β), simply because they explicitly concern our language while (γ) doesn’t.¹

To appeal to the Quinean criterion of ontological commitment is to make an argument of this form. In particular, it’s to make an argument of this form where p is existentially quantified: ‘Given that the truth-condition of “S” is “\(\exists x Fx\)”, if you accept that S, then you must accept that there are Fs’. Thus, my main argument that negative actions are events goes like this:

(1) ‘x NEG-φ-s’ is true \(\equiv\) \(\exists x\) NEG-φ-s.
(2) ‘x NEG-φ-s’ is true \(\equiv\) There exists an event that is x’s token NEG-φ-ing.
(3) ‘. . . x NEG-φ-s \(\equiv\) There exists an event that is x’s token NEG-φ-ing.

Regarding (1), I assume that the T-schema applies unproblematically to negative action sentences. It’s well known that the schema can’t apply in full generality, on pain of paradox (let ‘S’ be ‘S is false’). But all the solutions to this problem of which I’m aware aim to recover as many instances of the schema as possible. Since I’m not aware of any reason to think that applying the schema to negative action sentences leads to paradox, I assume it applies unproblematically to them. Point (2) is the Neo-Davidsonian claim I’ll be defending in detail in the central chapters of the book. Given (1) and (2), (3) follows unavoidably: an agent performs a negative action (i.e. she omits to do something, refrains from doing it, etc.) just in case there exists an event which is her token negative action (an omission, a refrainment, etc.).

Now, it doesn’t follow from (3) that there actually are such events as negative actions: the opponents of the Quinean criterion are right that we

¹ Thanks to David Liebesman for discussion.
can’t simply read the existence of Fs off the fact that some sentences in our language quantify over Fs. (As I note in Section 1.3.2, Quine himself was quite clear on this.) What follows from (3) is that if agents ever omit to do things, refrain from doing them, etc., then there are such as events as their token negative actions. You’re free to reject an ontology which includes such events if you find them problematic. But if you do, then you must deny that negative action sentences are ever true, and hence deny that agents ever perform negative actions.

As I said, I’ll defend this style of argument in more detail in Section 1.3.2. I turn now to sketching the argument of the book as a whole.

In Chapter 1, I lay out some preliminary doctrines regarding action and ontology, including an account of the ontological distinction between the things we do and our doings of those things, which will be important at various points throughout the book.

In Chapter 2, I discuss the phenomenon of negative action in more detail. I also provide a more detailed presentation of the motives behind each of (PNA 1)–(PNA 3), and show that the widespread acceptance of (PNA 3) can be traced to widespread acceptance of Deflationism about negative action sentences.

In Chapter 3, I consider and reject a popular solution to the problem, which is to reject (PNA 1) and insist that, while so-called negative actions are an important aspect of our agency, they nonetheless aren’t actions, strictly speaking. I argue that it’s more difficult than it might seem to draw the requisite distinction between actions and ‘mere manifestations of agency’, and to justify placing intentional omissions, refrainments, and the like in the latter category. If we want to do justice to the importance these behaviours have for our agency, we’re better off treating them as actions proper.

In Chapter 4, I turn my attention to (PNA 2), (PNA 3), and their linguistic motivations. Although Deflationism is widespread, both the thesis itself and the linguistic evidence for it are often left implicit or underdeveloped. In this chapter, I build an explicit case for Deflationism on behalf of its proponents. In brief, there’s a range of linguistic data which is typically used to motivate the Neo-Davidsonian approach to ordinary action sentences; however, the same kind of linguistic data undermines a simple Neo-Davidsonian approach to negative action sentences, and favours the Deflationist approach.

If we want to reject Deflationism, and with it (PNA 3), we need an alternative semantics for negative action sentences. I develop this alternative – which I call ‘the sophisticated Neo-Davidsonian approach’ – in
Chapter 5. On this approach, negative action sentences are existential quantifications over events which play what I call ‘the ensuring role’: to say that \( x \text{ NEG-} \varphi \text{-s at } t \) is to say that some behaviour of hers ensures (in a semi-technical sense, which I explain in detail) that \( x \) doesn’t \( \varphi \) at \( t \). Thus, to say that \( x \text{ NEG-} \varphi \text{-s} \) is to report the occurrence of a token NEG-\( \varphi \)-ing, now understood as an event which plays the ensuring role. This approach accommodates the linguistic data which scuttles the simpler Neo-Davidsonian approach discussed in Chapter 4.

By the end of Chapter 5, Deflationism and the Neo-Davidsonian approach are at a stalemate. In Chapter 6, I introduce a new range of linguistic data which my Neo-Davidsonian approach can accommodate but which the Deflationist approach can’t. Thus, taking all the linguistic evidence into account, my approach is to be preferred. We ought to think that negative action sentences are existential quantifications over ensuring-events, and hence (by the Quinean argument form sketched above) that if agents ever perform negative actions, there are such ensuring-events.

In Chapter 7, I turn to more robustly metaphysical issues. In my view, negative actions aren’t negative in a metaphysically deep sense. The ‘negativity’ of negative actions resides in the description of the ensuring role (\( x’s \text{ NEG-} \varphi \text{-ing at } t \) is whatever event ensures that no event is a \( \varphi \)-ing by \( x \) at \( t \)), while negative actions themselves are token-identical to ordinary, positive events. The reason is simple: negative actions are events which play the ensuring role; positive events play the ensuring role; therefore, negative actions are these positive events. This view is an analogue of ‘token realizer functionalism’ in the philosophy of mind, according to which a token mental state is whatever plays the functional role associated with that state (and hence, if that role is played by a token physical state, the mental and physical states are token-identical).

As I go on to show in Chapter 7, this realizer functionalist view is prima facie in conflict with the popular property-exemplification theory of events. Thus, it seems that if we want to endorse that theory of events, we must adopt a view on which negative actions aren’t identical to positive events, but are perhaps realized or constituted by them. However, I argue that we have independent reason to revise the theory in such a way that it’s consistent with realizer functionalism.

Finally, in Chapter 8, I consider a range of popular objections to the claim that negative actions are token identical to ordinary, positive events. Drawing on the details of my theory of negative actions, as well as the ontological framework defended in Chapter 1, I show that all these arguments fail.
In this book, I draw on ideas and arguments in the philosophy of action, the philosophy of language, and general metaphysics, in order to present a unified account of the nature of negative actions, our thought and talk about them, and their place in a theory of agency. On this account, negative actions can take their place alongside ordinary actions in event-based theories, and so we can retain those theories while recognizing the importance of negative actions to the fabric of human agency.
CHAPTER I
Action and Ontology

1.1 Introduction
In this chapter, I lay out some doctrines regarding action and ontology which will be important for understanding the problem of negative action and my solution to it. In Section 1.2, I explain the task of a metaphysical theory of action, draw a distinction between two kinds of action – the things we do and our particular doings of those things – and make clear what I will and won’t be assuming about the scope of a theory of action. In Section 1.3, I turn to the ontology of action, specifically the ontology of doings and things done. I defend the view that doings are events while things done are properties of agents, and defend commitment to entities of both kinds on Quinean grounds.

1.2 Action
1.2.1 The Task of a Theory of Action
Agents can behave in many different ways. I can raise my arms, kick my legs, run, jump, sing, and dance. I can also bump into doors, fall down stairs, and fall asleep. Some of these things aren’t like the others. Although all of these are things that I do, in some sense of the word ‘do’, there seems to be a more restricted sense of the word on which only some of them count as things I do, while the others are relegated to the category of ‘things that happen to me’. In some sense, raising my arms, kicking my legs, running, jumping, singing, and dancing are things I do, while bumping into doors, falling down stairs, and falling asleep are things that merely happen to me (Davidson 1971: 43). What’s the difference?

At its most abstract, the difference is between those things I do which are actions – i.e. those things in doing which I exercise my agency – and those which aren’t. The distinction between actions and non-actions is of
great importance for our understanding of ourselves and how we fit into the natural world. Agents, we think, have a kind of control over their behaviour that non-agents don’t, and which they exercise whenever they act.¹

The task of a theory of action is to explain the difference between action and ‘mere behaviour’. Not all ways of behaving can be classified one way or the other for all agents, or even for a single agent at all times. Breathing, for instance, is something I typically do automatically, and is not clearly a candidate for a behaviour through which I exercise my agency. However, I can take control of my breathing, and so breathing is sometimes a way of acting for me. Thus, action theorists focus on specific cases involving specific agents behaving in specific ways at specific times: the task of a theory of action is to explain the difference between cases in which agents exercise their agency and cases in which this isn’t so.

The task isn’t, in the first instance, to provide an extensional account of the distinction between cases of action and cases of non-action, i.e. an inventory of which cases are which. We already have (or think we have) a good sense of which cases are which (e.g. a case in which I raise my arm to signal a passing cab is a case of action; a case in which my arm rises as the result of an involuntary spasm isn’t), and a theory of action is to be judged, in part, by how well its dictates accord with our firmer judgments (e.g. a theory which says that I act when my arm rises due to a spasm but not when I raise my arm to hail a cab is in trouble). Rather than telling us which behaviours are actions and which aren’t, the task of a theory of action is primarily to tell us what the difference between action and mere behaviour consists in, and why any given case belongs in the category it does.

I say we have a good sense of which cases are which, not that we have a perfect sense. As in other areas of philosophy, disagreement is always possible about particular cases. Consider, for instance, alleged mental actions – roughly, actions which don’t involve an agent moving, positioning, or otherwise affecting those parts of her body that are visible to an outside observer. It certainly seems that some of the mental things I do are exercises of agency while others aren’t: making a decision seems to be something I do, while remembering seems to be something that merely...

¹ Bishop (1989: ch. 1), Shepherd (2014a), and Steward (2012a: chs. 1–3) explicitly tie agency to control over one’s behaviour. Bishop and Steward both stress the importance of control to our conception of ourselves, not only as agents, but as free and responsible ones. On the connection between agency, freedom, and responsibility, see also the introduction to Moya (1990).