



## Free Will and Modern Science

Richard Swinburne (ed.)

<https://doi.org/10.5871/bacad/9780197264898.001.0001>

Published: 2011

Online ISBN: 9780191754074

Print ISBN: 9780197264898

### CHAPTER

## 4 Dualism and the determination of action

RICHARD SWINBURNE

<https://doi.org/10.5871/bacad/9780197264898.003.0005> Pages 63–83

Published: December 2011

### Abstract

This chapter argues that it is most unlikely that neuroscientists will ever be able to predict human actions resulting from difficult moral decisions with any high degree of probable success. That result leaves open the possibility that humans sometimes decide which actions to perform, without their decisions being predetermined by prior causes. The chapter begins with two assumptions, which provide a different framework within which to work out how far human actions are predictable from that of Frank Jackson, and which lead to a different kind of conclusion.

**Keywords:** [neuroscientists](#), [human action](#), [moral decisions](#)

**Subject:** [Philosophy of Science](#), [Philosophy of Mind](#)

I argue in this chapter that it is most unlikely that neuroscientists will ever be able to predict human actions resulting from difficult moral decisions with any high degree of probable success. That result leaves open the possibility that humans sometimes decide which actions to perform, without their decisions being predetermined by prior causes. I need to begin with two assumptions, which provide a different framework within which to work out how far human actions are predictable from that of Frank Jackson (see the previous chapter), and which lead to a different kind of conclusion. I have space here only to provide brief justifications of these assumptions; for fuller justifications I must refer readers to other writings of mine.

# 1. Brain events and mental events interact

My first assumption (not held by Frank Jackson) is that there are goings-on (unchanging states or changes of states) of two non-overlapping kinds, ones which are public (i.e. equally accessible to all), and ones to which their subject has privileged access. I shall call these goings-on 'events'; the former I shall call 'physical events' and the latter 'mental events'. Physical events include brain events; anyone can discover as well as can anyone else what is going on in my brain. But my having a headache is an event to which I have privileged access, and so it is a mental event. Someone else can learn about my headache from my behaviour and from studying my brain (in the sense that studying these can show them that it's quite probable that I have a headache); but I also could learn about my headache in these ways, and yet I have a further means of learning about it by actually experiencing the headache. Some mental events do however have physical events as a constituent part. My seeing my desk is a mental event, but its occurrence entails that there is a desk present (a physical event). I define a pure mental event as one which does not entail the occurrence of a physical event—that is, it is not part of what is meant by the claim that that event occurs that some physical event occurs, although it is compatible with the claim that that event occurs that it is caused by a physical event. My headache is a pure mental event. (In another terminology pure mental events are 'narrow content' mental events.) All my subsequent references to mental events are to be understood as references to pure mental events.<sup>2</sup>

Among such mental events, some are such that necessarily if and when we have them we are (at least to some degree) conscious (aware) that we are having them. This group of mental events includes not merely sensations such as pains (often called 'qualia'), but also (occurrent) thoughts (which may be entertained without being believed). If I am not in any way aware that the thought 'today is Saturday' is now 'crossing my mind', it isn't crossing my mind. But my definition includes as mental events also events which still exist while the subject is not conscious of them, but of which the subject may become conscious from time to time. In this group are desires (inclinations to do some action to which the subject may or may not yield) and beliefs. I can desire to get home in time for lunch, or to write a great book—while not thinking about this or doing anything to achieve my desire. I have at any time lots of beliefs—about history or geography for example—of which I am not in any way conscious at that time. My concern in this chapter with intentions (that is, purposes) is solely with the intentions in our present intentional actions, for example my intention in walking along a certain road being to walk to the railway station, not with intentions to be executed later; and I shall understand by an 'intentional action' an action which one knows one is doing and means to do. In that sense of 'intention' each of us is always to some degree conscious (that is, aware) of our intentions. When we are to some degree conscious of any of our beliefs, intentions, and desires then (like sensations and thoughts), they are conscious events. And we are to some extent conscious of many such mental events all the time. For example all perceptions (and we perceive things all the time we are awake) involve not merely (or primarily) having sensations but consciously acquiring beliefs.

My second assumption is that—despite the recent work of neuroscientists pursuing the programme pioneered by Benjamin Libet and discussed (in their chapters in this volume) by Patrick Haggard and Tim Bayne—not merely are many conscious events caused by brain events, but conscious events often cause brain events and other conscious events. (Frank Jackson agrees that 'mental events' cause brain events, but he does not understand 'mental' events in the same way as I do.) My reason in brief for assuming that Libet-type experiments do not show that intentions do not in general cause the brain events which cause the bodily movements involved in intentional actions is that the evidence adduced by neuroscientists includes (as evidence necessary for establishing their conclusion) evidence about when subjects form various intentions (e.g. to move a hand). This evidence comes from what subjects tell the scientists about when they believe that they formed some intention. But scientists would only be justified in believing what the subjects tell them if they believed that subjects say what they do because they intend to tell the truth

p. 66

about their beliefs; that is, the scientists must believe that a subject's intention (together with a belief) causes him to say what he does. If they thought that the words coming out of a subject's mouth were caused only by a sequence of brain events themselves not caused by any intention to tell the truth, they would have no justification for believing what a subject tells them. From that it follows that, while some experimental results might show that sometimes intentions do not cause brain events, reaching that conclusion requires the assumption that often (e.g. when subjects tell scientists what they believe) intentions do cause brain events (and so bodily movements). Most of our beliefs about the world (everything most of us ever learnt about history or geography or science, including all the experimental evidence adduced by scientists and the scientific theories based on it) are derived from what other people tell us, and we believe what they say because we believe that an intention to tell the truth about what they believe causes ↪ them to say what they do. On the assumption that we are right to believe many of the things which we are told, we must assume that our informants' intentions, together with the beliefs implied by what they say, cause brain events which cause bodily movements.<sup>3</sup>

Intentions to perform basic actions (ones which we do, not by doing any other action), such as moving a hand or uttering a sentence, can no doubt cause effects without needing to be combined with many, if any, other conscious events (such as conscious beliefs) in order to do so. But most intentions are intentions to perform non-basic actions; a non-basic action is an action which an agent does by doing some other action—for example when I walk to the railway station by walking along a certain road. An intention to perform a non-basic action (e.g. to walk to the railway station) needs to be combined with a belief about which basic actions (e.g. of walking along a certain road) will lead to that intention being fulfilled.

As well as assuming that many conscious events cause brain events, I also assume that conscious events sometimes cause other conscious events in the kind of rational way codified by theorists of practical and theoretical reasoning, as when I perform a series of calculations leading to a belief about the result of a complicated sum, or when my belief that the coin has landed heads 99 out of 100 times causes my belief that it will land heads next time. Not merely does this seem fairly evidently so, but without this assumption we would have no justification for believing any of our inferences. If I am to be justified in believing some conclusion which I have reached by considering some argument, I must believe that I am moved (that is, caused) to believe it by reflecting on the earlier stages of the argument. Without this assumption we would have to think of ourselves as incapable of reaching conclusions (e.g. that some scientific theory is true) on the basis of evidence.

p. 67

So when we consciously form an intention (i.e. make a decision), we are often influenced by conscious beliefs (including value beliefs) and conscious desires. Value beliefs (as I shall understand this notion) are beliefs about the overall objective (including moral) goodness of doing different actions. In so far as I believe an action good to do I have a reason and so some inclination to form the intention to do it. While other beliefs need to be combined with some desire or prior intention in order to motivate us to act, value beliefs as such motivate us.<sup>4</sup> If I believe that it is obligatory to ↪ keep a promise, I will have some inclination to keep the promise; I couldn't think of it as obligatory if I did not. And the better I believe some action to be, the greater as such is my inclination to do it. But a value inclination may be weak, and I may yield instead to some other inclination.

A desire to do some action, as I shall understand this notion, is an inclination to form an intention to do the action when the subject believes that possible, independently of any inclination caused by a belief that it is objectively good to do it. Desires in this sense are a matter of what we 'feel like' doing. Beliefs and desires are, at a given time, involuntary states; I cannot change them at will. I believe that today is Saturday, that I am now in Oxford, that Aquinas lived in the thirteenth century, and so on and so on. I cannot suddenly decide to believe that today is Monday, that I am now in Italy, or that Aquinas lived in the eighteenth century. (Although I cannot change a belief at will, I can set about investigating a topic (e.g. when Aquinas lived), which might (but might not) lead to a change of belief.) Likewise we find ourselves desiring sleep or

food, fame or fortune; we cannot (normally) change these desires suddenly, but we can decide not to do the action which they incline us to do, and we can take steps which may lead to a change of desire in the future.

If I have no desire to do one available action rather than another (e.g. to give money to this charity rather than that charity), I form the intention to do what I believe is the best action to do, that is the one which, it seems to me, I have most reason to do. If I have no belief about the relative value of certain actions (e.g. to lunch at this restaurant rather than that one), I form the intention to do that available action which I most desire to do, that is the one which I find myself most inclined to do. But when I have equally strong desires to do each of two incompatible actions (and no stronger desire to do a different incompatible action) and no relevant value belief, or a belief that two incompatible actions would be equally good (and no other incompatible action would be better) and no relevant desire, and—above all—when what I most desire to do is incompatible with what I believe to be the best action to do, which intention I will form cannot be determined solely by the relative strengths of my desires or my beliefs about which action would be best to do. If I am caused to decide (as opposed to deciding without being caused to decide) to form an intention in such circumstances, the final outcome must be determined by brain events in a non-rational way.

p. 68 I have argued that beliefs and desires are caused, and I shall assume that all other mental events (conscious or not conscious) with the possible exception of intentions are also caused. Clearly some desires and sensations are caused directly by brain events without help from any other mental events; desires to drink or sleep, and sensations of pain or noise are surely in this category. But most of our desires, and—I suggest—all our beliefs and occurrent thoughts couldn't be had without coming in mutually sustaining packages of other beliefs and desires; or be conscious without being sustained by other conscious beliefs and desires. I could not desire to be prime minister without this desire being sustained by many beliefs about what prime ministers do, as well no doubt as some brain events causing me to desire to be famous or powerful. And I couldn't even come consciously to believe (through perceiving it) that there is a book on the table in front of me without having many other conscious beliefs, such as a belief that books are written by authors for people to read, and a belief that tables have flat surfaces, and so on.

In order to simplify the discussion, I shall assume that even if some mental (including conscious) events need other mental (including conscious) events to sustain them, a subject's total conscious state (all his conscious events) at a given time (with the possible exception of his intentions) is caused ultimately (sometimes via earlier conscious or other mental events) by his total brain state; and also that every type of total conscious state (with the same possible exception) is correlated with some type (or disjunction of types) of overall brain state. (So when a conscious event causes another conscious event, either the former or the latter causes a brain event correlated with the latter.) Most total conscious states will be large ones, full of beliefs and sensations (consider the many perceptual beliefs involved in coming to see a scene), and often some occurrent thoughts, desires or intentions. The brain state which is correlated with a conscious state will also normally be a large state; recent neuroscience suggests that it consists in a 'temporal synchrony between the firing of neurons located even in widely separated regions of the brain', between which there are 'reciprocal long-distance connections', a synchrony which attains a 'sufficient degree and duration of self-sustained activity'.<sup>5</sup> Different overall conscious states are correlated with different variants of this pattern of activity. So if we are to make predictions of future conscious events and brain events, we would need a theory of which aspects of a total brain state (which types of individual brain events) cause or are caused by which aspects of a total mental (including conscious) state. Then we could predict that any new total brain state which contained a certain type of brain event would cause a certain type of conscious event, such as a certain type of intention; or conclude instead that some intentions occur uncaused.

p. 69

## 2. Obstacles to assembling data for a mind-brain theory

To construct such a mind-brain theory we would need a lot of data in the form of a very long list of particular (what philosophers call 'token') conscious (and other mental) events occurring simultaneously with token brain events. To get information about which conscious events are occurring, we depend on the reports of subjects about their own conscious events. There are however two major obstacles which make it difficult or impossible to get full information from subjects.

The first obstacle concerns the 'propositional' mental events, occurrent thoughts, desires, beliefs, and intentions. I call them 'propositional events'<sup>6</sup> because they involve an attitude to a proposition (which forms the content of the attitude). A belief is a belief that such-and-such a proposition is true; a desire is a desire that such-and-such a proposition be true, and so on. The problem is that while the content of most of these events can be described in a public language, its words are often understood in slightly different senses by different speakers. One person's thought which he describes as the occurrent thought that scientists are 'narrow-minded', or the belief that there is a 'table' in the next room, has a slightly different content from another person's thought or belief, described in the same way. What one person thinks of as 'narrow-minded' another person doesn't, and some of us count any surfaces with legs as 'tables' whereas others discriminate between desks, sideboards, and tables.<sup>7</sup> This obstacle can be overcome, by questioning subjects about exactly what they mean by certain words. But it has the consequence that, since beliefs etc. are the beliefs they are in virtue of the way their owners think of them, far fewer people have exactly the same particular beliefs, desires, etc. as anyone else than one might initially suppose—which makes the kind of experimental repetition which scientists require to establish their theories much harder to obtain. And it seems most unlikely that any two humans understand all their words in exactly the same way, and so have exactly the same concepts as each other.

There is however a much larger obstacle to understanding what people tell us about their sensations. This is that we can understand what they say only on the assumption that the sensations of anyone else are the same as we would ourselves have in the same circumstances—and that is often a highly dubious assumption. This obstacle applies to all experiences of colour, sound, taste and smell (the 'secondary qualities'). We can recognize when someone makes the same discriminations as we do between the public properties of colour etc., but we cannot check whether they make the discriminations on the basis of the same sensations as we do. While it might seem counter-intuitive to suppose that green things look to one group of people just like red things look to another group of people, while red things look to the first group just like green things look to the second group, other possibilities seem less counter-intuitive. Maybe green things look a little redder to some people than they do to others,<sup>8</sup> or coloured things look fainter to some people than to others, when neither of these differences affect their abilities to make the same discriminations.

We could rule out such possibilities on grounds of simplicity (that it is simpler and so more probable to suppose that these things do not happen), if it were the case that which neurons have to fire in which sequence at which rate in order to produce a sensation which subjects call 'green' (or whatever) were exactly the same in all humans. But in view of the differences between the brains of different humans, that seems very improbable. It's much more likely that sometimes for two different people different neurons produce a sensation which they both call 'green'. The different reactions which people often have to the same input from the senses supports the hypothesis that the sensations caused thereby are sometimes different in different people. Some people like the taste of curry, others don't. There are two possible hypotheses to explain this: curry tastes the same to everyone but some people like and some people don't like this taste, or curry tastes differently to different people. It would seem highly arbitrary to suppose that the first explanation is correct—let alone suppose that a similar explanation applies to all different reactions to tastes.

I need however to make a qualification to all this, that while we may be unable to understand the natures of the individual sensations of others, their sensations may exhibit patterns which are the same as some publicly exemplifiable patterns (of primary qualities such as shape). Thus a mental image of a square has the same shape as a public square. The lines which make up the image may have peculiarities of colour which the subject cannot convey, but he can convey the shape. I shall return to this point later.

### 3. Obstacles to forming a predictive theory from the data

So, bearing in mind these limits to the kinds of mental data we can have, what are the prospects for forming a theory supported by evidence which will not merely explain and so predict how brain events cause sensations, beliefs, and desires, but how these (together with brain events) cause our subsequent intentions?

What makes a scientific theory such as a theory of mechanics able to explain a diverse set of mechanical phenomena is that the laws of mechanics all deal with the same sort of thing—material objects—and concern only a few of their properties—their mass, shape, size, and position—which differ from each other in measurable ways (one has twice as much mass as another, or is three times as long as another). Because the values of these measurable properties are affected only by the values of a few other such properties, we can have a few general laws which relate two or more such measured properties in all objects by a mathematical formula. We do not merely have to say that, when an inelastic object of 100g mass and 10m/sec velocity collides with an inelastic object of 200g mass and 5m/sec velocity, such and such results, with unconnected formulas for the results of collisions of innumerable inelastic objects of different masses and velocities. We can have a general formula, a law saying that for every pair of inelastic material objects in collision the quantity of the sum of the mass of the first multiplied by its velocity plus the mass of the second multiplied by its velocity is always conserved. But that can hold only if mass and velocity can be measured on scales—for example, of grams and metres per second. And we can extend mechanics to a general physics including a few more measurable quantities (charge, spin, colour charge, etc.) which interact with mechanical quantities, to construct a theory which makes testable predictions.

p. 72

A mind-brain theory however would need to deal with things of very different kinds. Brain events differ from each other in the chemical elements involved in them (which in turn differ from each other in measurable ways) and in the velocity and direction of the transmission of electric charge. But mental events do not have any of these properties. The propositional events (beliefs, desires, etc.) are what they are, and have the influence they do in virtue of their propositional content, often expressible in language but a language which—I noted earlier—has a content and rules differing slightly for each person. (And note that while the meaning of a public sentence is a matter of how the words of the language are used, the content of a propositional event such as a thought is intrinsic to it; it has the content it does, however the subject or others may use words on other occasions.) Propositional events have relations of deductive logic to each other; and some of those deductive relations determine the identity of the propositional event. My belief that all men are mortal wouldn't be that belief if I also believed that Socrates was an immortal man; and my thought that ' $2=1+1$ , and  $3=2+1$ , and  $4=3+1$ ' wouldn't be the thought normally expressed by those equations if I denied that it followed from them that ' $2+2=4$ '. And so generally, much of the content of the mental life cannot be described except in terms of the content of propositional events; and that cannot be done except by some language (slightly different for each person) with semantic and syntactic features somewhat analogous to those of a public language. The rules of a language which relate the concepts of that language to each other cannot be captured by a few 'laws of language' because the deductive relations between sentences and so the propositions which they express are so complicated that it needs all the rules contained in a dictionary and grammar of the language to express them. These rules are independent rules and do not follow from a few more general rules. Consider how few of the words which occur in a dictionary

can be defined adequately by other words in the dictionary, and so the same must hold for the concepts which they express; and consider in how many different ways describable by the grammar of the language words can be put together so as to form sentences with different kinds of meaning, and so the same must hold for the propositions which they express.

p. 73 So any mind-brain theory which sought to explain how prior brain events cause the beliefs, desires, etc. which they do would consist of laws relating brain events with numerically measurable values of transmission  $\hookrightarrow$  of electric charge in various circuits, to conscious (and non-conscious) beliefs, desires, intentions, etc. with a content individuated by sentences of a language (varying slightly for each person), and also sensations. The contents of mental events do not differ from each other in any numerically measurable way, nor do they have any intrinsic order (except in the respect that some contain others—e.g. the belief about the book contains a belief about its uses). Those concepts which are not designated by words fully defined by other words—and that is most concepts—are not functions of each other. And they can be combined in innumerable different ways which are not functions of each other, to form the propositions which are the contents of thoughts, intentions, etc. So it looks as if the best we could hope for is an enormously long list of separate laws (differing slightly for each person) relating brain events and mental events without these laws being derivable from a few more general laws.<sup>9</sup>

p. 74 Could we not at least have an ‘atomic’ theory which would relate particular brain events involving only a few neurons to particular aspects of a conscious state—particular beliefs, occurrent thoughts, etc., the content of which was describable by a single sentence (of a given subject’s language), in such a way that we could at least predict that a belief with exactly the same content would be formed when the same few neurons fired again in the same sequence at the same rate (if ever that happened)? The ‘language of thought’ hypothesis<sup>10</sup> (LOT), which takes seriously the analogy of the brain to a computer which manipulates symbols, seems to involve some version of an atomic theory. It claims that there are rules relating brain events and beliefs of these kinds, albeit a very large and complicated set of them. It holds that different concepts and different logical relations which they can have to each other are correlated with different features in the brain. For example, it holds that there are features of the brain which are correlated with the concepts of ‘man’, ‘mortal’, and ‘Socrates’, and that there is a relation R which these features can sometimes  $\hookrightarrow$  have to each other. When someone believes that Socrates is mortal, R holds in their brain between the ‘Socrates’-feature, and the ‘mortal’-feature; when someone believes that Socrates is a man, R holds between the ‘Socrates’-feature, and the ‘man’-feature; and when someone believes that all men are mortal, R holds between the ‘man’-feature and the ‘mortal’-feature. (The holding of this relation might perhaps consist in the features being connected by some regular pattern of signals between them.) The main argument given for LOT is that unless our brain worked like this, the operation of the brain couldn’t explain how we reason from ‘all men are mortal’ and ‘Socrates is a man’ to ‘Socrates is mortal’, since our reasoning depends on our ability to recognize the relevant concepts as separate concepts connected in a certain particular way. Beliefs, and so presumably other propositional events, the theory claims, correspond to ‘sentences in the head’.

I argued earlier however that no belief can be held without being sustained by certain other beliefs—for logical reasons; which other beliefs a given belief is thought of as entailing determines in part which belief the latter belief is. Now consider two beliefs, whose content is expressed in English by ‘this is square’ and ‘this has four sides’; someone couldn’t hold the first belief without holding the second. So these two beliefs cannot always be correlated with different brain events, since in that case a neuroscientist could eliminate the brain event correlated to the latter belief without eliminating the brain event correlated to the former belief. On the other hand these two beliefs cannot always be correlated with the same brain event since someone can have the belief ‘this has four sides’ without having the belief ‘this is square’. It follows that propositional events are correlated with more than one (type of) brain event. That leads to the view that propositional events only occur as part of a total mental state, including many other mental events, and it is this whole mental state which is correlated with a whole brain state without there being correlations



between separate parts of the mental and brain states. This view is that of connectionism,<sup>11</sup> the rival theory to LOT. Mind-brain relations are holistic. Only if connectionists hold, as they often do, that mental events are identical with (or supervene logically on) individual brain events, is it an objection to connectionism that brain events do not have a structure corresponding to that of a human language. But given my initial assumption ↪ that mental events are events distinct from brain events, mental events can have a sentential structure without brain events having this. So, given connectionism, a mind-brain theory could at best only predict the occurrence of some conscious event in the context of a large mental state (consisting of many beliefs, desires, etc., some of them conscious) and of a large brain state (events involving vast numbers of neurons).

p. 75

We have seen that we must suppose that mental events often cause other mental events in a rational way. As I illustrated earlier, the laws of rational thought include the criteria of valid deductive inference, and—since the validity of an inference between sentences depends on which propositions are expressed by which sentences, and that depends on the meanings and arrangements of the words the sentences contain—these can only be stated fully by lists as long as those of the dictionary and grammar of a human language. These laws also include the criteria of cogent inductive inference (that is, criteria of inductive probability, of which propositions make which other propositions probable). They also include the criteria for forming value beliefs. But each human person has slightly different criteria of these kinds. Further, of course, humans do not always follow their own criteria of rational thought, and so we would need laws stating when and how brain events disturb rational processes. These latter laws would vary with the overall mental and brain states of the subject, and the mental states which disturb rationality (such as beliefs) would often need to be described in terms of the concepts with which that subject operates (e.g. some particular fixation preventing someone reasoning rationally about a particular subject matter).

Further, insofar as mental events cause other mental events in a rational way, the influence of mental events depends on their strength; and (apart from occurrent thoughts) they all have different strengths. One person's sensation of the taste of curry is stronger than another person's. One person's belief that humans are causing global warming is stronger than another person's (that is, the first person believes this proposition to be more probable than does the second person). Yet while subjects can sometimes put sensations in order of strength in virtue of their subjective experience, what they cannot do—despite 150 years of work on 'psychophysics'—is to ascribe to them numerical degrees of strength in any objective way.<sup>12</sup> (How do I answer the doctor who asks 'Is this pain more ↪ than twice as severe as that pain?') And, despite 80 years of work on 'subjective probability', the same applies to beliefs and other propositional events.<sup>13</sup> Such differences affect behaviour in a rational way. Someone is more likely (despite counter-influences) to stop eating curry, the stronger is the taste of curry which she dislikes. Someone is more likely (despite counter-influences) to choose to travel by bus rather than by car because of a belief that humans are causing global warming and a belief that it is good to prevent this, the stronger are those beliefs. In order to measure the influence of sensations, beliefs, etc. on intentions (in a situation where there are many conflicting influences), we need a measure of their absolute strength which can play its role in an equation connecting these; and subjects cannot provide that from introspection. Neuroscience might discover that greater frequency of certain kinds of brain event causes the beliefs caused by those brain events to be stronger. But for prediction of their effects we'd need to know how much stronger were the resultant beliefs. So we'd need a theory by means of which to calculate this, which gave results compatible with subjects' reports about the relative strengths of their beliefs. But the brain circuits, rates of firing, etc., which sustain beliefs in different subjects are so different from each other that it is difficult to see how there could be a general formula connecting some feature of brain events with the strength of the mental events which they sustain. So the most we could get is a long list of the kinds of brain activity which increase or decrease the strength of which kinds of mental events.

p. 76



p. 77 So the part of a mind-brain theory which predicts human intentions and so human actions would consist of an enormous number of particular  $\hookrightarrow$  laws relating brain events to subsequent sensations, thoughts, beliefs, and desires (some of them conscious), and these (together with other brain events) to subsequent intentions, having this kind of shape:

Brain events ( $B_1, B_2 \dots B_j$ ) + sensations ( $M_1 \dots M_e$ ) + Thoughts ( $M_f \dots M_i$ ) + Beliefs (including value beliefs) ( $M_j \dots M_k$ ) + Desires ( $M_k \dots M_l$ )  $\rightarrow$  Intention ( $M_n$ ) + Beliefs (about how to execute the intention) ( $M_p \dots M_q$ ) + Brain events  $\rightarrow$  bodily movements.

The B's describe events in individual neurons, and each law would involve large numbers of these; the M's describe particular mental events with a content describable by a short sentence, and with a strength. The strength of an intention measures how hard the agent will try to do the intended action.

In cases where mental events alone determine the resulting intention, we can no doubt often predict that intention in virtue of the general principles outlined at the beginning of the chapter—e.g. where there is a strongest desire and no relevant value beliefs, or a strongest value belief and no relevant contrary desire, and the agent believes that he is able to do the action, the agent will form the intention to act on the strongest desire or value belief (even if we cannot predict whether the intention will be strong enough to be executed). But where brain events interact with mental events to form desires and beliefs and thereby to determine subsequent intentions, the previous argument has the consequence that—if determined—the outcome would be determined for each person by one of an enormous number of different laws relating total brain states to total mental states, including large total conscious states. So we could not work out what a person will do on one occasion when she had one set of brain events, beliefs and desires, on the basis of what she (or someone else) did on a previous occasion when she had a different set differing only in respect of one relevant belief. For there would be no general rule about the effect of just that one change of belief on different belief and desire sets; the effect of the change would be different according to what was the earlier set, and what were the brain states correlated with it. But no human being ever has the same total brain state and mental state at any two times or the same total brain state and mental state as any other human does at any time, and—I suggest—no human being considering a difficult moral decision ever has the same conscious state, let alone the same brain state in the respects which give rise to consciousness and determine its transitions, as at another time or as any other human ever. For making a difficult moral decision involves  $\hookrightarrow$  taking into account many different conflicting beliefs and desires. The believed circumstances of each such decision will be different, and (consciously or unconsciously) an agent will be much influenced by her previous moral reflections and decisions.

p. 78

Consider someone deciding how to vote at a national election. She will have beliefs about the moral worth of the different policies of each party, and the probability of each party executing its policies; she will desire to vote for this candidate and against that candidate for various reasons (liking or disliking them for different reasons); she will desire to vote in the same way as (or in a different way from) her parents, and so on. But each voter will have slightly different beliefs and desires of these kinds. So because each voter's total conscious state would never have occurred previously, there could not be any evidence supporting a component law of the mind-brain theory to predict what would happen this time. A very similar conscious state might have occurred previously (in the same or a different voter), supporting a detailed law about the effects of that similar conscious state. But that suggested law (weakly supported by one piece of evidence) would (because of the slight difference in the conscious state) only predict what would happen this time with a degree of probability surely less than a half. Add to this the point that the part of the overall brain state which determines the strength of the different events constituting the conscious state, and how rationally subjects will react to them, will almost certainly be different on the different occasions. Add to all this the points made in Section 2 about the difficulty involved in getting some of the evidence required to support any mind-brain theory, and I conclude that it is most unlikely that a prediction about which

difficult moral decision someone would make, and so which resulting action they would do, could ever be supported by enough evidence to make it probably true. Human brains and human mental life are just too complex for humans to understand completely.

That conclusion has a crucial consequence that those brain events which cause the movements which constitute human actions will never be totally predictable. But even if it should turn out that the behaviour of other physical systems is totally predictable, it should not be too surprising that the brains of humans (and perhaps higher animals) are different, since the brain is unlike any other physical system in that it causes innumerable nonphysical events.

## 4. What neuroscience can discover

The limits to the ability of neuroscience to predict arise from the enormously large number of detailed laws which would have to govern any interaction of many different kinds of mental (including conscious) events and brain events. But neuroscience may be able to discover, and has begun to discover, mind-brain laws which do not involve such complicated interactions. Thus it has begun to discover which particular brain events are necessary and sufficient for the occurrence of those non-propositional events which do not involve the inaccessible aspects of sensations, but only the patterns of sensations. A mental image has the same sort of properties of shape and size as the properties of public objects such as brain events. So neuroscience is on the way to discovering a law-like formula by which it can predict from a subject's brain events both the images caused by the public objects at which she is looking and the images which she is intentionally causing.<sup>14</sup> But that formula will not tell us what the subject regards their image as an image of —e.g. as an image of a television set or of a shiny box. Which beliefs subjects acquire about what they are seeing is clearly going to vary with their prior beliefs about the way objects of different kinds look, e.g. that something of such and such a shape is a television set. But if the neuroscientist discovers these prior beliefs in some other way (e.g. from what subjects tell him, or by analogy with his own beliefs), then he should be able to predict from a subject's brain events not merely the shape of the image which she is having, but the subject's belief about what she believes that she is seeing.

Similar considerations apply to the other senses. Which words a subject hears depends on the pattern of sensed sounds rather than their intrinsic qualities; and patterns of sensed sounds have the same describable shape as patterns of public sounds. So it should be possible to construct a formula describing how the brain events caused by certain patterns of public sounds cause patterns of sensed sounds. Given people's linguistic beliefs (their beliefs about what words mean) discoverable in some other way, it should then be possible to predict from their brain events what they understand to be the content of what is being said to them. So scientists should be able to arrange for sentences to be 'heard' by the deaf whose auditory nerves

↳ no longer function, by means of electrodes in their brain causing the appropriate brain events.

Desires to do basic actions can occur in the absence of a large set of beliefs. Hence neuroscience can discover the brain events which are the immediate causes of desires to form intentions to do basic actions, such as to drink. It can also discover the brain events which are the immediate effects of intentions to perform basic actions (e.g. move a hand, or utter a certain word), these being intentions which can be had independently of any other beliefs. That will enable it to detect what 'locked in' people are trying to do, and so set up some apparatus which will enable them to succeed.<sup>15</sup> But in order to predict which non-basic action a subject has the intention of performing, a neuroscientist would need to know the subject's beliefs about which basic actions would bring about the performance of the non-basic action. Hence we need to know subjects' linguistic beliefs in order to know which proposition as opposed to which words (defined by the sounds which constitute its utterance) subjects are trying to utter.

Neuroscience may be able to make various kinds of statistical predictions, to the effect that a change in the pattern of certain kinds of brain events will probably lead to an increase or decrease in the strength of certain kinds of desire or belief and so to the probability of certain intentions. Thus it may be able to discover how certain brain events affect the relative strengths of very general kinds of desire (e.g. for fame or power). Desires influence but, when the subject also has competing particular desires and value beliefs, do not determine a subject's intentions and so behaviour. And which intention a general desire will tend to cause will depend on the subject's beliefs (e.g. about how fame can be obtained). So again in the absence of a formula for calculating beliefs of any complexity from brain events, and in the absence of a formula for calculating intentions from competing beliefs and desires (and brain events), all we can hope for is statistical predictions to the effect that the more or less of some physical quantity that brain events have, the greater or less the desire to do so-and- so, and so—probably—the greater the proportion of subjects who will do so-and-so. Hence drugs or mirror neurons may indeed promote or diminish altruistic desires,<sup>16</sup> or strengthen or weaken a desire to commit suicide. ↵ But such increases or decreases of desires only yield statistics; they don't tell you who will do what, since we all have different rival desires of different strengths and different value beliefs of different strengths.

It follows however, finally, that neuroscience should be able to predict what individual humans will do in order to execute certain general instructions which have as a consequence that their behaviour must depend on only one simple desire of a kind caused directly by a brain event. For example, in the Libet (2004: ch. 4) experiments discussed earlier in this volume subjects were told to move their hand at any time within a short period when they decided to do so; and since they would not have had any value beliefs about when to do so, they must have decided to do so when they 'felt like' it, i.e. desired to do it. Such a desire is like an itch and so presumably has a direct cause in a brain event. So in this case neuroscience may be able to correlate prior brain events with the movements which they cause, via the desire to cause them. If however subjects disobeyed the instructions, and didn't move their hand within the period—either because they didn't feel the requisite desire or because they had rival desires (e.g. to be a nuisance) or value beliefs (e.g. that it was immoral to take part in the experiment), their actions would not count in assessing the experiment. So 100% success in predicting hand movements under these experimental conditions is by no means impossible. But once again that tells us nothing about how people will behave in situations of conflicting desires and value beliefs.

So despite the possibility (and in some cases the actuality) of all these advances in neuroscience, the main point of this chapter remains, that for the prediction of individual behaviour in circumstances where there are different variables, both brain events and mental events of different and competing kinds and strengths affecting the outcome, neuroscience would need a general formula well supported by evidence to enable it to relate the strengths of these kinds of events to each other; and that cannot be had.

## 5. Intentions are probably undetermined

---

p. 82

Yet, even if it is unpredictable which intention we will form in such circumstances and how strong it will prove, what reason do we have for supposing that that intention (with its particular strength) is not caused (in a way too complicated to predict) by brain events? After all, I have acknowledged, our intentions are often caused—when they are caused by  $\hookrightarrow$  a strongest desire and we have no contrary value belief, a strongest value belief and we have no contrary desire, and when our desires and value beliefs are in opposition to each other. My answer is that it is just under the circumstances where desires or value beliefs are of equal strength or in opposition to each other, that we are conscious of deciding between competing alternatives. We believe that it is then up to us what to do, and we make a decision. Otherwise, as we realize, we just move along habitual paths. It is a basic principle of rationality that things are probably the way they seem to be (in the sense that we are inclined to believe that they are) in the absence of counter-evidence. All science begins from experience, and experience is experience of the way things seem to be (in the physical or mental world). When we make a decision, it seems that we choose and are not caused to choose as we do; in other cases it does not seem that we are making a choice which is up to us there and then. So, in the absence of counter-evidence (in the form of a causal theory of our behaviour in such circumstances, rendered probably true by much evidence), when we make a decision, we are probably doing so without being caused to do so.

## References

---

Davidson, D. (1980) Mental events. In *Essays on Actions and Events*, Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Fodor, J.A. (1979) *The Language of Thought*, Cambridge, MA: Harvard University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Gray, J. (2004) *Consciousness: Creeping up on the Hard Problem*, Oxford: Oxford University Press.

Kay, K.N. *et al.* (2008) Identifying natural images from human brain activity, *Nature*, 452: 352–5. [10.1038/nature06713](#)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Kellis, S. *et al.* (2010) Decoding spoken words using local field potentials recorded from the cortical surface, *Journal of Neural Engineering*, 7: 1–10.

[Google Scholar](#) [WorldCat](#)

Laming, D.R.J. (2004) Psychophysics. In Richard L. Gregory (ed.) *The Oxford Companion to the Mind*, Second edition, Oxford: Oxford University Press, pp. 771–3.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Libet, B. (2004) *Mind Time*, Cambridge, MA: Harvard University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Lycan, W.G. and Prinz J.J. (eds) (2008) *Mind and Cognition: An Anthology*, Third edition, Oxford: Blackwell.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Robinson, R. (2009) Exploring the ‘global workspace’ of consciousness, *PLoS Biology* 7(3): doi10.1371/journal.pbio.1000066.

[Google Scholar](#) [WorldCat](#)

Shinkareva, S.V. *et al.* (2008) Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings, *PLoS ONE* 3(1): e1394.doi10.1371/journal.pone.0001394.

[Google Scholar](#) [WorldCat](#)

p. 83 Swinburne, R. (1997) *The Evolution of the Soul*, Revised edition, Oxford: Oxford University Press. [10.1093/0198236980.001.0001](#)

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#) [Crossref](#)

Swinburne, R. (1998) *Providence and the Problem of Evil*, Oxford: Oxford University Press. [10.1093/0198237987.001.0001](#)

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#) [Crossref](#)

Swinburne, R. (2007) From mental/physical identity to substance dualism. In P. van Inwagen and D. Zimmerman (eds) *Persons: Human and Divine*, Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Swinburne, R. (2011) Could anyone justifiably believe epiphenomenalism?, *Journal of Consciousness Studies*, 18(3/4): 196–216.

[Google Scholar](#) [WorldCat](#)

Zak, Paul J. *et al.* (2009) Testosterone administration decreases generosity in the ultimatum game, *PLoS ONE*, 4(12): e8330.doi:10.1371/journal.pone.0008330.

[Google Scholar](#) [WorldCat](#)

## Notes

---

- 1 Many thanks to Daniel Robinson and to two anonymous reviewers for very useful comments on an earlier draft of this paper.
- 2 For a full justification of the (to many of us) obvious point that there are these two kinds of 'going-on', see Swinburne (1997: Part I). This first assumption is the assumption of property (and so event) dualism, which is the moderate kind of dualism. The more radical kind of dualism, substance dualism, which I also advocate but do not assume in this chapter, is discussed in Howard Robinson's chapter in this volume. For the superiority of the way of making the mental/physical distinction in terms of a subject's privileged access, over other ways, see Swinburne (2007: 142–4 and nn. 3 and 4).
- 3 For full argument in justification of this claim see Swinburne (2011).
- 4 For argument in support of this point, see Swinburne (1998: Additional note 3).
- 5 Gray (2004: 173 and 175). The 'global workspace' model has been confirmed by recent work of Raphael Gaillard and others; see Robinson (2009).
- 6 They are sometimes called 'intentional' events, but I avoid this label since it leads to confusion between intentions and the wider class of 'intentional events'.
- 7 Though it hardly needs such support, this point is borne out by recent experiments showing that presenting some image to a group of subjects produced in all subjects similar patterns of activity in different regions, but slightly different patterns for each subject. See Shinkareva *et al.* (2008).
- 8 Even if two groups of subjects typically agree in the percentage of 'redness' shown by greenish colour samples, that won't show that the 'pure green' or 'pure red' samples look the same to both groups, and so that a '10% red' sample looks the same to both groups.
- 9 Donald Davidson is well known for arguing that 'there are no strict psychophysical laws' (1980: 222). This thesis is stronger than mine, but his reasons for it are similar to mine. However he uses this thesis in defence of his theory of 'anomalous monism', that all events are physical, some events are also mental, and so physical-mental causal interaction (which we both recognize) is law-like causal interaction of two physical events. But, contrary to Davidson, I am assuming (for reasons stated briefly at the beginning of this chapter) that there are events of two distinct types, physical and mental; and so I reject Davidson's resulting theory.
- 10 Originally put forward by Fodor (1979).
- 11 For a selection of papers on both sides of the language-of-thought/connectionism debate see Parts II and III of Lycan and Prinz (2008).
- 12 See Laming (2004): 'Most people have no idea what "half as loud" means. In conclusion, there is no way to measure sensation that is distinct from measurement of the physical stimulus.' The latter sentence is a bit pessimistic; it might be possible to measure it from the (in some sense) strength of the brain event which caused it. But there is no reason to suppose that a noise which was by such a measure half as loud as a second noise would seem that way to its hearers.
- 13 There is a long tradition beginning with the work of F.P. Ramsey, of attempting to measure someone's degree of belief in a proposition (the 'subjective probability' which they ascribe to it) by the lowest odds at which they believe that they would be prepared to bet that it was true. If someone is, they believe, prepared to bet £N that *q* is true at odds of 3-1 (i.e. they would win £3N if *q* turned out true, but lose their £N if *q* turned out false) but not at any lower odds (e.g. 2-1), that—it was claimed—showed that they ascribe to *q* a probability of 1/4 (because then in their view what they would win multiplied by the probability of their winning would equal what they would lose multiplied by the probability of their losing). But that method of assessing subjective probability will give different answers varying with the amount to be bet—someone might be willing to bet £10 at 3-1 but £100 only at odds of more than 4-1; and people have all sorts of reasons for betting or not betting other than to win money.
- 14 K.N. Kay *et al.* (2008) devised a decoding method which made it possible to identify, 'from a large net of completely novel natural images, which specific image was seen by an observer'.
- 15 See the work (described in Kellis *et al.* 2010) being done to detect the brain events caused by intentions to utter certain sounds, which will enable computers to translate these into speech, when people are 'locked in' and unable to communicate by speech.
- 16 Paul J. Zak *et al.* (2009) found that increasing testosterone in men makes them less generous in the game situations created by psychologists.