FREE WILL AND MODERN SCIENCE

EDITED BY
Richard Swinburne
Contents

Foreword: questions of freedom  vii  PETER SIMONS
List of figures  xvi
List of contributors  xvii

Introduction: plan of the volume
RICHARD SWINBURNE  1

1 Does brain science change our view of free will? 7  PATRICK HAGGARD

2 Libet and the case for free will scepticism 25  TIM BAYNE

3 Physicalism and the determination of action 47  FRANK JACKSON

4 Dualism and the determination of action 63  RICHARD SWINBURNE

5 On determinacy or its absence in the brain 84  HARALD AMTANSPACHER AND STEFAN ROTTER

6 G"odel's incompleteness theorems, free will and mathematical thought 102  SOLOMON FEFERMAN

7 Feferman on G"odel and free will: a response to Chapter 6 123  J.R. LUCAS
Contents

Foreword: questions of freedom .......................... vii
PETER SIMONS
List of figures ........................................... xvi
List of contributors ....................................... xvii
Introduction: plan of the volume ........................ 1
RICHARD SWINBURNE

1 Does brain science change our view of free will? .......... 7
PATRICK HAGGARD

2 Libet and the case for free will scepticism ............... 25
TIM BAYNE

3 Physicalism and the determination of action ............. 47
FRANK JACKSON

4 Dualism and the determination of action .................. 63
RICHARD SWINBURNE

5 On determinacy or its absence in the brain ............... 84
HARALD ATMANSCHACHER AND STEFAN ROTTER

6 Gödel’s incompleteness theorems, free will and mathematical thought .............................. 102
SOLOMON FEFERMAN

7 Feferman on Gödel and free will: a response to Chapter 6 ................. 123
J.R. LUCAS
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 The impossibility of ultimate responsibility?</td>
<td>126</td>
</tr>
<tr>
<td>GALEN STRAWSON</td>
<td></td>
</tr>
<tr>
<td>9 Moral responsibility and the concept of agency</td>
<td>141</td>
</tr>
<tr>
<td>HELEN STEWARD</td>
<td></td>
</tr>
<tr>
<td>10 Substance dualism and its rationale</td>
<td>158</td>
</tr>
<tr>
<td>HOWARD ROBINSON</td>
<td></td>
</tr>
<tr>
<td>11 What kind of responsibility must criminal law presuppose?</td>
<td>178</td>
</tr>
<tr>
<td>R.A. DUFF</td>
<td></td>
</tr>
</tbody>
</table>

*Guide to background reading*  

*Index of names*  

---

Foreword

**Question**

PETER SIMON

The problem of responsibility for criminal acts has resisted solution since the ancient world. The question, however, from the perspective of modern jurisprudence, is how period, to today, is a complex one. The issue, more than any informed community, is what does it mean to be responsible for the dispute. Is responsibility, as we have come to understand it, convinced, when we act, that we know the end we have in mind? Endowed, rather, with the will we do, and the absence of our will. What then is responsibility and what is the role of the will? It is a given. It is a moral concept.

The traditional approach to responsibility has been to identify there a specific concept of responsibility, the name or description of which we may not be able to point to in practice. This concept is describable as a complex of mental and social phenomena, as a psychology. It refers to particular descriptions, or to phenomena.

Any number of different descriptions of responsibility may be set in opposition to one another and, indeed, to Napoleon’s dictum: “When the army has to drop atomic bombs, they are not to be used with choices of national survival.” Whether or not we agree where to go from here...
Figures

1.1 Cortical structures involved in generation of voluntary action. 11
1.2 Transcranial magnetic stimulation over the motor cortex can be used to cause involuntary movements of the hand. 13
1.3 Schematic representation of the Libet experiment. 17
1.4 The design of Wenke and colleagues' experiment. 21

Contributors

Harald Atmanspacher is Head of the Department of Theory and Data Analysis, Institute for Frontier Areas of Psychology, Freiburg.

Tim Bayne is University Lecturer in the Philosophy of Mind at the University of Oxford and a Fellow of St Catherine's College.

R.A. Duff is Emeritus Professor of Philosophy, University of Stirling, Professor of Law, University of Minnesota, and a Fellow of the British Academy.

Solomon Feferman is Emeritus Professor of Mathematics and Philosophy, Stanford University.

Patrick Haggard is Professor of Cognitive Neuroscience, University College London.

Frank Jackson holds fractional research appointments at the Australian National University and La Trobe University, and is Visiting Professor of Philosophy, Princeton University, and a Corresponding Fellow of the British Academy.

J.R. Lucas was a tutorial Fellow of Merton College, Oxford, and is a Fellow of the British Academy.

Howard Robinson is University Professor, Department of Philosophy, Central European University, Budapest.

Stefan Rotter is Professor of Computational Neuroscience, Faculty of Biology, University of Freiberg, and Bernstein Centre, Freiberg.

Peter Simons is Professor of Philosophy, Trinity College, Dublin, and a Fellow of the British Academy.

Helen Steward is Senior Lecturer in Philosophy, University of Leeds.

Galen Strawson is Professor of Philosophy, University of Reading.
3

Physicalism and the determination of action

FRANK JACKSON

1. What I'll do

There is no single version of physicalism. There is no single argument for
physicalism. There is, accordingly, no standard answer concerning the
implications of physicalism for the causation of human action by mental
states. Of necessity, I will be highly selective. I will start by describing the
version of physicalism I favour and saying why I favour it. We will then
discuss what it implies about the connection between subjects' mental
states and what they do, and thereby for the determination and pre-
dictability of our actions. This serves as a precursor for a short discussion
of the implications of physicalism for the possibility of free action. Near the
end I will say something about a version of physicalism that I don't favour
but which many like. This version is commonly thought to have very
different implications for the determination and prediction of action from
the version of physicalism that I like. We will see that this is indeed the case
but not in ways that help much when we are worrying about the impli-
cations of physicalism for the possibility of free action.

2. The preferred version of physicalism

Give a stone a kick and not much happens. Give a person a kick and quite
a lot may happen. Present an apple with a quadratic equation and not much
happens. Present a person with a quadratic equation and you may get the
answer. Present a person with a set of house plans and you may end up
with a house, no point though in doing this with a tree.
These commonplaces remind us that persons determine functions from inputs to outputs of a highly sophisticated kind. What is more, we know these functions connect closely with a person’s mental states. If someone is required to cross a minefield, their response tells you very quickly where they believe the mines are located. That is, the function you observe from minefield to response tells you something about their mental states, namely, where they believe the mines are located. Likewise, we know that the function from house plans to built houses goes via what the potential builder sees and wants. Showing plans to someone who can’t see them or doesn’t want to build the house won’t get you very far.

What’s the take-home message? That mental states play functional roles. This is hardly news. It is, surely, something the folk know. It is not something discovered by high-powered cognitive science. The folk have known since the dawn of time about the functional impairments that go along with failing eyesight and hearing, for example, and that these impairments relate directly to the nature of their visual and auditory experiences.

There is a simple way to go from these commonplaces to a version of physicalism. Here is how it goes:

Premise 1
Mental state so and so = the state that plays role such and such. (Commonplace we have just been talking about)

Premise 2
The state that plays role such and such = brain state whatever, having the kinds of properties that figure in current and future neuroscience. (Empirical discovery)

Conclusion
Mental state so and so = brain state whatever. (By the transitivity of identity)

Some say the first premise, suitably expanded and detailed, is a conceptual truth of a kind with ‘Poisons make people sick’. The idea is that for each mental state there is a definition functional role. This issue doesn’t matter for our purposes right now (although we return to the question later). What matters for now is that something like the first premise is true for the mental states we will be concerned with, among which are our intentions and our decisions to do this or that.

Some say the conclusion isn’t really an identity; it should be read as a claim about constitution, analogous to the relation between a table and the molecules that compose it. Mental states are constituted by, while not being identical with, brain states. This issue is by the way for us here.

Some say the conclusion is a contingent truth; others say that – for suitable values of ‘such and such’ and ‘whatever’ – it is a necessary a-posteriori truth. We will return to this issue later.

What is important at this stage is the commitment to mental states playing causal roles. The functional roles we are talking about require that mental states play causal roles. They may require more than this, and what that more might be is contentious, but they require at least that mental states play causal roles. This means that Richard Swinburne (this volume) and I agree that epiphenomenalism is false. We are at one in rejecting the idea that mental states do nothing.

3. Mental states directed at mental states

We are now in a position to say something quickly about the debate initiated by the work of Benjamin Libet; the debate that in part prompted the symposium that this volume arises from. (For details of Libet’s results, see Haggard, this volume.)

Among the functional roles played by mental states are roles directed towards mental states themselves. Not all of the functional roles concern how we relate to our surroundings. For example, perceptual experience is a putative response to the environment, a response that represents that things are thus and so, whereas the belief that one has an experience that represents that things are thus and so is a response to the perceptual experience itself. Similarly, although the job of implementing one’s intention to obey someone is to cause behaviour in conformity with that intention, the job of one’s awareness of implementing one’s intention to obey someone is to let one know that you are indeed implementing the intention – something it can be very useful to know. Likewise, the functional role of the belief that it is raining is to record the putative information that it is raining and to lead to behaviour that tends to serve one’s desires if that belief is true – taking an umbrella before going out, say. But that isn’t the role of the awareness that one believes that it is raining. The role of the awareness that one believes that it is raining is, for example, to make it explicit to one what one believes and to generate, in animals with language, sentences like ‘It is raining’.

This means – at least for anyone who likes the functionalist approach to mental states – that whenever we consider how best to interpret the results of experiments in psychology that draw heavily on subjects’ own reports
 Physicalism and the determination of action

of, or judgements about, the mental states they are in, including, as it might be, the time at which the mental states occur, we need to be alert to the possibility that what's being reported is a property of the awareness of the mental state rather than the mental state putatively under investigation.

Now what Libet did was to devise an ingenious way of timing when subjects take it that they are making a certain decision. He found that the time at which they take it that they are making the decision is somewhat later than the time at which the relevant changes in the brain – those that lead to behaviour that implements the decision – commence. But this doesn’t tell us that the decision itself takes place somewhat later, a result which would raise awkward questions about the causal role of the mental state of deciding to act. It tells us instead something about subjects’ awareness of the time at which they implement a decision.1

4. Why believe the preferred version of physicalism?

The argument set out earlier is valid. We have seen why we should accept the first premise. The question that remains, accordingly, is the case for the second premise. This has been, as one would expect, the subject of much debate. (See any recent text in the philosophy of mind.) But, in a nutshell, the case rests on neuroscientific optimism combined with a general hostility to the idea of outside causal influences on the physical world.

There are lots of things we don’t know about how our brains work but there doesn’t seem any particular reason to hold that we will find some kind of gap in the causal story about how stimuli at our sense organs lead to motor responses via happenings in the brain, a gap suggesting that some kind of causal influence from outside the physical world is at work in the brain, or at some point in the causal path from environmental input to behavioural response.

The last paragraph is essentially a reminder of the line of thought that is often encapsulated by saying that the physical world is causally closed. Although physicalists like myself are supporters of the thesis that the physical world is causally closed – read in a way that takes account of indeterminism – there is a well-known problem in saying what ‘physical’

means here. I’ll say something about this later. For now, we can think in terms of the kinds of properties neuroscientists appeal to when they explain the way the impact of the world on our sense organs translates into motor responses. This means that the core idea behind the second premise is that when neuroscientists detail and expand our understanding of the causal paths from environmental impacts on our peripheries to happenings inside us, especially happenings in our brains, and the pathways to and from our brains, and from happenings elsewhere inside us, to behavioural responses, they won’t find causal transitions that suggest the presence of outside causal influences.

Two more points to note about the second premise. First, I talked of the case for it resting on neuroscientific optimism. This suggests that anyone who wishes to resist the kind of physicalism I like only needs to be a bit of a pessimist. An easy thing to be, one might think. However it isn’t that simple. We want the words that come out of our mouths, the marks we make on paper and what we do to keyboards to carry information about our mental states. If that were not true, my diary entries for, say, 2001 would not be a good source of information about what I thought and felt back then, and emails from your friends would not be good sources of information about what they are thinking and feeling. Anyone who wants to hold that a critical part of our mental nature lies in properties additional to those posited in neuroscience and the functional roles they fill, must hold that, somehow or other, these extra properties stand in information-preserving causal relationships to the words we use to talk about them, the various marks that appear on paper and what we do to keyboards. They need to be optimistic – implausibly optimistic, say I – about how the laws that govern the connections between the extra properties and what happens in our brains ensure that the distal happenings at our motor extremities carry information about the extra properties.

Secondly, there is a feature of our version of physicalism that some philosophers worry about. (My sense is that philosophers worry about it more than brain scientists.) Our version is a version of internalism about mental states; it affirms that they are located inside our heads, because that is, as a matter of fact, where the crucial functional roles are played.2 We should, that is, locate them inside our heads for the same reason we should locate the memory processing in an iMac somewhere behind the screen.

1 Which plausibly isn’t the time of the awareness per se but the time at which the awareness represents the deciding to be taking place. Thanks here to Nic Shea.

2 The roles themselves, however, concern how things are outside the subject. Part of what makes some state a desire for coffee is its connection with coffee.
The worry is that the states we are talking about are types or kinds of state, and types and kinds as such aren’t located. (This seems, for example, to be Williamson’s concern (2009: 331).) They are abstract entities. It is their instances – the things that are of the relevant type or kind – that are located. This might suggest that the only sensible thing to mean by the location of a mental state is the location of someone who is in that state, and we aren’t located inside our heads.

There is nothing to worry about here, or so it seems to me. There are many states – many kinds of states – that a largish physical structure is in by virtue of some part of it being a certain way. A computer stores a memory in virtue of a part of it being the right way; a person has cancer in virtue of a part of that person being a bad way; someone has arthritis in virtue of a part of a joint being inflamed; and so it goes. In cases like these, although the largish thing is in the state, the state itself is located where the part of the big thing that is the relevant way is located.

You now have before you the version of physicalism I like; it is a version of the type-type mind-brain identity theory (see, e.g., Armstrong 1968). What does this version of physicalism say about the determination of action and the predictability of action, the questions that relate most directly to the topic of this volume?

The answers depend on two matters: how we should think about the first premise of the argument we started with, and how we should respond to what physics tells us about determinism.

5. The functionalist premise

There are broadly three ways we can think of the first premise: as a bit of common sense about mental states but nothing more than that; as something that, with suitable elaboration, can be turned into a conceptual truth; or as something that, with suitable elaboration, can be turned into an account of how we reference-fix on mental states.

The first downplays the importance of the functional roles played by mental states. I think it goes too far. People who suffer strokes want above all else to get function back. I think they would be surprised to be told that the connection between getting function back and getting, say, the ability to see things off to their left was purely accidental. Or consider what is involved in remembering something. The role, the functional role, of remembering in delivering information about the past seems much more than a well-known but accidental fact about remembering. Again, the observation that belief is a state designed to fit the world, and the observation that desire is a state designed to make the world fit it, seem to be much more than known truths about belief and desire. They seem to go to the heart of what belief and desire are. Finally, it is nothing more than an interesting truth that people in pain tend to behave in ways that they believe will minimize their pain.

Suppose then that the functional roles of mental states are more than interesting facts about them. One response, as we observed earlier, is to think of being in a mental state of such and such a kind as like being poisoned: something definable in terms of a certain functional role. In the case of being poisoned, we can state the functional role. In the case of any given mental state we cannot. We can give rough indications – our talk above of belief being a state designed to fit the world is an example, as is our talk earlier of perceptual experience being a putative response to the environment that represents that things are thus and so – but a statement with all the bells and whistles in place will be beyond us. Here is the point where those who like this approach, and I’m an example, talk about implicit theories. They say that mental states are defined by implicitly grasped functional roles.

The other response is to think of the functional roles played by mental states as reference fixers. It is obvious that we have a pretty good grasp of the functional roles played by various mental states. That is evident from the way we are able to explain and predict people’s behaviour in terms of their mental states, and the way we use how they respond to various situations to reach reliable judgements about their mental states. But we need not think of these remarks as support for the existence of some sort of conceptual connection between mental states and functional roles. Maybe the functional roles serve to fix the reference of the mental state terms. On this view, the essential nature of any given mental state is the brain state it is identical with, much as, according to many, the essential nature of water is H₂O. But our knowledge that certain functional roles are occupied preceded our knowledge of the brain. Something about the way we and our fellows interacted with the world led us to hypothesize that, inside each of us, there are, for example, states that carry information about the world and which interact with each other in systematic ways. How else are we able to avoid the holes in the ground and the tigers? And how else can we explain the fact that, over time, people get better at avoiding the holes in the ground and the tigers – obviously some sort of useful
information processing and updating is going on inside us. Accordingly, we should, says the reference-fixing view, think of the functional roles as alerting us to the existence of internal states that are connected one to another, and to the world around us, in various ways. These internal states are our mental states but their essential nature qua mental states isn’t given by their functional roles. That’s a contingent feature of them that led us to hypothesize their existence in the first place. Their essential natures are instead given by their a-posteriori-determined neurological natures.

We have here two very different ways of thinking about the identities between mental states and brain states that are the conclusion of the two-premise argument for physicalism we set out earlier. If the connection between functional role and mental state is conceptual, the mind-brain identities will be contingent a-posteriori truths. They will be like:

The poison most beloved of mystery writers = arsenic.

On the other hand, if the connection between functional role and mental state is a reference-fixing one, the identities between mental states and brain states will be necessary a-posteriori ones like (according to most philosophers and I won’t be arguing the toss here):

Water = H₂O

However, although this is a big difference, as far as the topic of this volume goes, I suspect it is a small difference. On both views, enough information about brain state and functional role will determine without remainder what subjects do (modulo indeterminacy, a matter we will address shortly), and will do so in a way that allows one to predict what subjects will do. Supporters of the two views will disagree about how particular parts of the body of information are crucial and why, and how the functional part relates to the neurological part, but they will agree on the point that if you tell me enough about John Doe’s internal functional roles and brain processes, you tell me what determines what he will do, and you do so in a way that allows me to predict what he will do.

Now I need to mention a complication that, I am glad to report, causes less trouble than one might have feared.

6. The distinction between action and movement

There is an important distinction between behaviour in the sense of what a person does, and behaviour in the sense of the movements of the person’s body (raw behaviour, as it is sometimes called). My reaching for a glass is something I do. In doing it, my arm will move through space in some way that brings my hand near to the glass. The action and the movement are not the same. Indeed, the same action can involve different movements: there are many ways of hailing a taxi, and the same movement can be part of different (intentional) actions: for example, turning a light on in Australia and turning a light off in America require the same downward movement. This invites the question, were we talking about actions or movements in the previous few paragraphs?

However, for the two kinds of physicalism we were discussing, it doesn’t matter – that’s why the complication doesn’t cause trouble. According to physicalists of either of these two kinds, enough detail about the causal history of the movements of a person’s body delivers automatically the causal history of the person’s actions. The reason is that, for these physicalists, each and every action is nothing over and above some movement of the body with the right causal history in terms of mental states. This means that, for them, the causal history of a body’s movements and the causal history of the actions of a person with that body are not two different topics. Although classifying by actions differs from classifying by movements, what a person does is determined without remainder by, and is fully predictable given enough information about, physical antecedents alone – or rather this is true given determinism. It is time to discuss the implications of indeterminism in physics.

7. The implications of indeterminacy

It is very likely that determinism is false⁴ but it turns out that this makes less difference than one might have expected for the implications of physicalism for the possibility of free action. There are two cases to consider. One is that, as far as mental states, actions and bodily movements go, determinism might as well be true. This

⁴ Although experts in quantum theory tell me that interpretations of quantum theory in which determinism is true are undergoing something of a revival.
is because, at the macroscopic level of actions, of those brain states which are mental states, and of bodily movements, in the vast majority of cases things proceed as if determinism were true. The indeterministic effects at the sub-microscopic level wash out as we aggregate. Obviously, in this case the truth of indeterminism is by the way in the debate over free will.

The other, more likely case (as I understand matters but I defer to experts in physics) is that there is significant indeterminacy at the level of actions, of those brain states which are mental states, and of bodily movements. In this case, what persons do at a time is not determined by, and is not fully predictable from, how things are beforehand with their brains and surroundings even if physicalism is true. However, although what they do today isn’t fully determined by or predictable from how things were beforehand in terms of their brains and functional states, the chances of their doing this, that or the next thing are fully determined by, and predictable from, how their brains are at a given earlier time. What they in fact do today then has an irreducibly chancy connection with those earlier chances. Although physicalists must deny that the way my brain plus surroundings are, here and now, fully determine everything I do in the future, they should affirm that the way my brain plus surroundings are, here and now, fully determine the chances for every possible future thing I do, and that the passage from a chance of my doing X to my subsequently doing X is irreducibly chancy.

A natural thought is that in this second case, the case where the indeterminacy at the sub-microscopic level doesn’t wash out as we aggregate, the truth of indeterminism has significant implications for whether or not physicalism poses a threat to the existence of free action. This, however, turns out to be a mistake. The reason goes back to the overall nature of the debate over free will and determinism.

In the centuries-long debate over whether the hypothesis that what we do at any given time is fully determined by, and predictable from, how things were before that time, does or does not imply that we never act freely, the warring parties divide into two camps. One camp, that of the compatibilists, says that determinism is compatible with acting freely on occasion. Their key argument is that determinism is consistent with an action’s being under an agent’s control, in the sense that the agent would have done otherwise if they had chosen to do otherwise. For, they note, determinism is compatible with the causal path from the past to an action having, as a crucial component, the agent’s decision. What is more, they observe, determinism is consistent with the action’s reflecting the agent’s

desires and character. For determinism is consistent with an action’s being caused by an agent’s desires and character. And, according to compatibilists, something like the foregoing two facts, or some suitable tweaking of them, is all that needs to be the case for an action to count as a free action. Now, all this remains true when we factor in chances. As far as the point about control is concerned, the only change is that our actions will be under our control in the sense that had we chosen otherwise, the chances would have been different, and, moreover, the chances will normally be very close to one or zero as the time of action approaches. As far as the point about the connection between, on the one hand, desire and character, and action, on the other, goes, it remains true that there is a distinction between actions that reflect one’s desires and character and those that don’t. Making the connection between action and character and desires to some extent chancy doesn’t alter this.

Those in the other camp – those who hold that determinism is incompatible with acting freely – insist that all this talk of control, of the causal path to action going via an agent’s decision, of an action’s possibly reflecting the agent’s character and desires, is focusing on the wrong question, is indeed a piece of misdirection, on the part of compatibilists. The challenge that determinism poses to the existence of free actions concerns the causal antecedents of a person’s decisions, desires and character, and the like. Those in the incompatibilist camp allow that determinism is consistent with some actions being such that had the agent decided to do otherwise, they would have done otherwise. So what, they say. The key question concerns the causal origins of the decisions agents in fact make. They are unlikely to find it comforting to be told that these origins include a chancy element.

Likewise, incompatibilists grant that some actions reflect the desires and characters of those who perform them but insist that the key question, when

4 See, e.g., Ayer (1954). Recent defenses of compatibilism – often influenced in one way or another by Frankfurt (1971) – have introduced significant modifications but they don’t affect our point here and below. The modifications respond to the point that human beings can stand in complex relationships to their desires, their decisions and their characters. For example, human beings can have desires and characters they would prefer not to have, or ones they judge to be, in some sense, irrational, or ones that they, in some sense, disown. Many smokers arguably fall into one or more of these categories. This suggests that compatibilists should specify what it is to act freely in terms of acting in accord with some suitably restricted set of desires, decisions and characters – for example, the ones that agents themselves want to have. But none of this affects the point that whether or not there is an element of chance is by the way in the debate.
our topic is the existence of free actions, is the causal origins of the desires and characters. They worry that our desires and our character come from factors outside our control – our genetic make-up and the conditions into which we were born, for example. But discovering that the connection between our genes and our desires and character is in part chancy isn’t the discovery that we have control over our genes or the economic circumstances of our parents. In sum, incompatibilists urge that the challenge posed by determinism to the existence of free action lies in what determinism says about the causal antecedents of an agent’s decisions, character and desires, and the like; the crucial point being that those causal antecedents are outside an agent’s control. (For a version of this line of thought, see Strawson, this volume.) Bringing chance into the picture doesn’t somehow make these past factors into ones that are under the agent’s control.

I am saying nothing about which party, the compatibilist or the incompatibilist, wins the debate. Our point is that it is hard to see how inserting chances into the picture could affect who wins, one way or the other.

8. Anomalous physicalism

We have been discussing the implications of a version of physicalism that puts identities between mental states and brain states centre stage. There is, however, a very different version of physicalism, a version that holds it is a mistake in principle to identify the mental and the physical, in the sense of identifying mental and physical *kinds*. At first blush, this kind of physicalism might seem good news for those who worry about the implications of physicalism for freedom. I’ll be saying that the good news is not that good.

This kind of physicalism starts from the notion of a physical property, relation, law and entity, where they are of a kind with those to be found in physics, chemistry and biology. But this isn’t a definition of the physical. It is an *indication* of the kinds of properties etc. intended. The official definition of the physical is simply that it is whatever we need to give a full account of the non-sentient. The thesis of physicalism is then the claim that our world is nothing over and above a huge aggregation of elements that are themselves non-sentient. The sentient emerges from this aggregation in the same way that a house emerges when you put the various bits, none of which is itself a house, together in the right way. We know that aggregation creates new patterns. That’s the fun in playing with Lego. The idea, in the case of the mind, is that, when we suitably aggregate the non-sentient, some of the new patterns emerging from the aggregation will be mental properties. Indeed, the idea is that exactly this happened when each and every one of us was conceived. Conception triggered a complex process of aggregation of bits that are, in themselves, non-sentient to deliver over time something that is sentient.

(This definition makes a potentially controversial assumption about the foundations of quantum mechanics. The assumption is that talk of the observer in quantum mechanics is a case of bad labelling. Anyone who reads about the two-slit experiment will be struck by the thought that there is something deeply puzzling going on and should hope that smart physicists will sort the mess out for us. One possible sorting out will give the observer *as such* an important, fundamental role. Would this mean that sentence is a fundamental ingredient of our world? That seems unlikely. Despite all the uncertainty about what to say about the two-slit experiment, it seems that the role of the observer is to be a measuring device, something big enough to carry information about what is going on. Sentence as such isn’t part of the story.)

So far we have not said anything that necessarily goes against an identity version of physicalism – the identity version can agree that we are nothing over and above aggregations of the physical, in the sense of physical we have been discussing. The clash comes when we add the distinctive thesis of anomalous physicalism – or, rather, the distinctive thesis of a ‘vanilla’ version of anomalous physicalism. Our interest is in the core idea that underpins a range of views about the relation between the physical and the psychological that agree in repudiating dualism while insisting that the kind of reductionist picture that comes from functionalism or the type-type identity theory is mistaken. (For versions of this core idea, see Davidson 1980a, 1980b, and the references in Davies 2000, and especially his discussion of ‘upwards explanatory gaps’; for some dissent from the core idea, see Jackson 2010.) The distinctive thesis is that there are no patterned dependencies running from the physical to the mental. What exactly does this come to?

Physicalism is committed to the supervenience of the mental on the physical. If the mental is nothing over and above a complex aggregation of elements that are purely physical, then duplication in the physical implies duplication in the mental. This means there exists a raft of true conditionals of the form

If X is in \( P \), then X is in \( M \).

\[ \text{If } X \text{ is in } P, \text{ then } X \text{ is in } M. \]
which take us from a full enough specification of X's physical nature to the mental state that supervenes on that physical nature.

However, this is consistent with the absence of a patterned dependence of the mental on the physical. Supervenience per se only requires that, for each physical antecedent, there is exactly one mental state. It does not require that, for each mental state, the physical antecedents of the conditionals with that mental state as consequent have any degree of unity. The claim of anomalous physicalism is that if we group the conditionals together by sameness of mental consequent, we won't find any pattern uniting the physical antecedent that is sufficient to allow us to identify mental types with physical types.

How might one argue for anomalous physicalism? We cannot argue in the standard way from the problems for epiphenomenalism about mental properties. We cannot, for example, offer an argument of the following form:

Pain is causally efficacious with respect to behaviour (denial of epiphenomenalism)

All causally efficacious properties with respect to behaviour are physical properties (the causal closure of the physical)

Therefore, pain is a physical property (by transitivity)

Anomalous physicalism precisely denies that there are any true identities between mental and physical properties. We have to appeal instead to a difference principle. It says that the mental makes a difference. Had there been no mental properties, things would have been very different behaviourally. This claim is consistent with denying that mental properties are causally efficacious. Take the property of being fragile. Plausibly, though contestably, being fragile is not a causally efficacious property. What makes a dropped glass break is not its being fragile but is instead the categorical basis of its fragility—the thinness and internal structure of the glass, as it might be. Maybe mental properties are like that; indeed many say exactly this. All the same, being fragile may make a difference in the following sense: it may well be true that had the glass not been fragile, it would not have broken. Those who hold that being fragile isn't causally efficacious don't go around telling people that it is fine to knock fragile glasses over.

How do we use the difference principle to argue for physicalism? By arguing that we know enough about what the world is like and how it works, to be confident that the only way things might have been different

in the relevant behavioural ways is if they were different physically. Thus, if being different mentally makes for behavioural differences, this can only be because any and every given mental difference corresponds precisely to some physical difference. But then physical sameness implies mental sameness. What is the best explanation of this striking fact? The thesis that the mental emerges from the aggregation of physical.

9. Does anomalous physicalism change things that much?

A natural first thought is that the whole issue looks very different if one embraces anomalous physicalism. No longer does one have the spectre of patterns in the physical imposing themselves on the mental, and especially on those aspects of the mental especially concerned with intentional action. In some sense, we get a genuine autonomy of the mental and the intentional but in a way that avoids implausible metaphysical 'additions' of the kind that make so many uncomfortable with dualism. And it is worth recalling that this general idea has a substantial history that predates modern discussions of physicalism (see, e.g., Ryle 1954).

Here's the problem; the bad news. We noted earlier the distinction between movement and action. For identity versions of physicalism the distinction isn't of great moment as far as the prediction of intentional action goes. This is because, as we observed, full information about physical antecedents of movements delivers full information about actions according to identity versions of physicalism. This, however, is not the case if anomalous physicalism is true. Although the physical determines the mental, if anomalous physicalism is true, it will often be the case that we cannot, as a matter of principle, infer the mental from the physical. We will have, as we might say it, determination without predictable determination of the mental by the physical. In this sense, the mental will be autonomous. In this sense, we have some 'wriggle' room that might suggest a way of making sense of free action. This is because an agent's intentional actions, what the agent does, won't be fully predictable going by the physical alone, and the same goes for the chances of the agent's doing this or that. However, and this is where the bad news comes in, it remains true that full information about physical antecedents delivers full information about movements and full information about the subsequent distribution of physical objects. But if you worry that the predictable determination of what you do, or the chances of what you do, at any given time, by a purely physical past would
rob you of freedom of action, it would seem ad hoc not to worry that the predictable determination of the movements of your body, or of the location of objects in your vicinity, or the chances of same, by a purely physical past robs you of freedom of movement or of the freedom to rearrange how things are around you. If your movements aren’t free, what freedom of action remains? It would be strange to celebrate how free one’s actions are, while conceding that you aren’t free to cause your eyelid to close, or to change the angle your arm is at by some given number of degrees, or to change the location of a chess piece on a chess board. A gift of freedom of intentional action that comes without freedom of movement or freedom to change things around one would not be much of a present.

10. The upshot

What does physicalism imply about the determination of action? We have seen that there is no standard answer. It all depends on the correct version of physicalism, and on the implications of indeterminism in physics. However, perhaps surprisingly, we have also seen that physicalism has no particular implications for the existence of free action, or so I have argued.

References


4

Dualism and the determination of action

RICHARD SWINBURNER

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

1 Many thanks to Daniel Robinson and to two anonymous reviewers for very useful comments on an earlier draft of this paper.