



inely novel or distinctive features of reality, with their own causal powers. (Sometimes emergentism is defined in terms of the unpredictability or inexplicability of mental properties or states from knowledge of the physical states on which they supervene. I once expressed things this way; but I now prefer to define the doctrine to be formulated in purely metaphysical terms.) The question then is whether emergentism is compatible with the causal closure of the physical; this depends on what the causal closure thesis is, and in section 18 of EM I discuss one form in which the thesis might be denied.

The problem of consciousness is the other horn of the dilemma which constitutes today's mind-body problem. Consciousness is a problem in at least two ways for physicalism. First, it presents an explanatory challenge: how can physicalists explain what it is for something to be conscious in purely physical terms? This is the famous 'explanatory gap' problem. Second, it presents an ontological challenge: if physicalism is true then there has to be a necessary connection between mental and physical entities. But it seems conceivable that mental and physical entities could exist separately; therefore physicalism is false. This line of thought is sometimes expressed today as the famous 'zombie' argument. The response to these arguments given in sections 26–29 of EM is, in effect, to accept them and show why they are not problematic on an emergentist conception of mental phenomena. An emergentist can accept that there are limits to the physical explicability of mental phenomena, and indeed I argue that a physicalist should accept this too: this is the proper lesson of the 'knowledge argument' against physicalism. But an emergentist, unlike a physicalist, can also accept the conclusion of the zombie argument too, since part of the point of emergentism is that the dependence of the mental upon the physical is a matter of lawlike dependence rather than metaphysical necessity. Emergentism, although undeniably controversial, provides a consistent solution to the problem of mental causation and the problem of consciousness.

One of the lessons of chapter 3 of EM is that settling the debate between physicalists and non-physicalists settles very few other problems in the philosophy of mind. We understand very little about thought, consciousness, perception and action even if we have established—for example, by the general kind of causal argument discussed in chapter 2—that they must be physical. Hence there is more to the philosophy of mind than the debate over physicalism. The rest of EM is concerned with this; and this brings us to its major theme, intentionality.

I claim that intentionality is the mark of the mental, by which I mean that all mental phenomena exhibit intentionality. (In EM I do not defend the converse claim, that only mental phenomena exhibit intentionality, even though I do believe it.) Intentionality is the same thing as a state of mind's having an intentional object: something on which it is directed. It is crucial that 'object' is not interpreted as meaning 'thing' or 'entity'. An intentional object is not an entity of a certain kind. It cannot be, since some intentional objects do not exist. Yet all entities exist. In other words,

to talk about intentional objects is to talk about that on which one's mind is directed, whether or not it exists. I take it for granted that our minds can be directed on the non-existent, although this is what gives rise to some of the hardest problems of intentionality. A conclusion I draw from this fact is that intentional states cannot, in general, be relations to their objects.

To say that all states of mind must have an intentional object, then, is just to say that it is impossible for there to be a state of mind which is not *about* something, which is not directed on something. There are however different ways in which a state of mind may be directed on something: wanting something, disliking it and merely contemplating it are all intentional states. The way in which they differ is not in their object, but in what I call the intentional *mode*. (The intentional mode is what Husserl in the *Logical Investigations* called the intentional *quality*; other philosophers, who think that all intentional states are propositional attitudes, would call it the *attitude*.)

Intentional states can, however, be identical in mode and intentional object, but nonetheless differ. This is because they may differ in the way in which they present their object—or, as I shall say, in the *aspect* under which they present it. This kind of difference in intentionality I describe as a difference in intentional *content*. For a state to have intentional content is for it to have an (existing or non-existing) intentional object presented under a certain aspect. Since it is impossible, I argue, for an intentional state to have an object without presenting it under some aspect, then it follows that all intentional states have intentional content. I do not say that the intentional content of a state of mind is the way the world is represented as being, since some intentional states (e.g. desires, hopes) do not represent the world as actually *being* a certain way, but rather represent a non-actual condition of the world. Nor do I say that all content is propositional—that is, assessable as true or false—since there are many states of mind (notably object-directed emotions like love and hate) which do not have propositional contents. Many intentional states do have propositional content—these are the propositional attitudes.

I therefore understand intentionality in terms of the three central ideas of intentional object (where object is not understood as *thing* or *entity*), intentional mode (belief, desire, hope, fear etc.) and intentional content (that which characterises that on which the state is directed, and the aspectual shape of that presentation). Chapters 3, 4 and 5 apply these ideas to the paradigm mental phenomena of consciousness, thought and perception.

Those who reject intentionalism sometimes do so because they think that there are paradigmatic mental states which are not intentional. The two supposed types of example are bodily sensations like pains, and so-called 'undirected' emotions or moods. In chapter 3 of EM I discuss the case of sensations. (I do not yet have a developed view on the nature of emotions and moods, but I hope to be able to say something useful about them in the future.) I reject the view that bodily sensations are 'pure

qualia’—properties of a person which altogether lack intentionality. My reason for this is that sensations have a necessary felt location within the body, and this is a form of directedness. The location in the body is the intentional object of the state or event of feeling the sensation. Some intentionalists claim that pains represent damage to the body; I reject this as a phenomenological claim—a headache, for example, need not connote damage to the body in any sense—and argue instead that the phenomenology of pains and other sensations are characterised by their mode and their content, and the latter is their location in the body apprehended under a certain aspect.

Chapter 4 is concerned with the intentionality of thought: that is, of the propositional attitudes and conscious episodes of thinking. I argue that beliefs are never conscious, though we can become conscious of what it is that we believe, and that not all intentional mental states are propositional attitudes. The bulk of the chapter is concerned to defend the coherence of an internalist account of thought and the attitudes, contrary to the prevailing externalist orthodoxy. I am persuaded to defend internalism partly because I see the possible non-existence of objects of thoughts as a central and characterising feature of intentionality, partly because I believe in the coherence (though of course not the truth) of sceptical hypotheses like the brain-in-a-vat, and partly because I am unpersuaded by the positive arguments for externalism. I distinguish the question of externalism from the question of ‘*de re*’ thought. Although there are *de re* ascriptions of thoughts, the idea of *de re* thought as such is incompatible with one of the core claims about intentionality expounded in chapter 1: that every intentional state has an aspectual shape. But externalism need not be the thesis that there are *de re* thoughts in this sense. In its most plausible version, it is rather the thesis that there are fully intentional states of mind—with aspectual shape—which essentially depend on the existence of their objects. I do not argue directly for internalism, but instead dispute the most influential ‘Twin Earth’ style of argument for externalism, and provide a sketch of an internalist account of demonstrative thought.

The final part of EM’s investigation of intentionality concerns perception. Chapter 5 concentrates on two main areas of the philosophy of perception: (i) the traditional problem of perception—posed by the ‘argument from illusion’—and how an intentional conception of perception is the solution to this problem; and (ii) the phenomenal character of perception, in particular the question of whether perception involves the awareness of qualia (understood as ‘pure qualia’ in the sense defined above). I deny that there are any qualia in perception, and hold that the phenomenal character of perception can be fully explained in terms of intentional mode and content. The content of perception has, however, a special character: it is *nonconceptual*, in the sense that being in states with that content does not require possession of the concepts which are required for having other attitudes with that content. This special character is part of what explains the phenomenal character of perception.

In what follows I offer some comments on the criticism and discussion of aspects of this book put forward by Ana Gavran, Dunja Jutonrić, Olga Markić, Nenad Mišćević and Ksenija Puškarić. These papers were originally presented at the 2002 Dubrovnik workshop on Mental Phenomena. I am very grateful to Ana, Dunja, Olga, Nenad and Ksenija for the careful attention they have paid to my work, for the useful discussions in Dubrovnik, and for giving me this opportunity to clarify some of the ideas contained in my book.<sup>2</sup>

### *Reply to Ana Gavran*

Ana Gavran is sceptical of my criticism of the argument for an externalist theory of mental content. Externalists say that the intentional contents of a thinker’s thoughts is not determined by the local or intrinsic properties of the thinker; internalists deny this. Hence externalists deny the principle of local supervenience of the intentional on the physical: it is not the case, according to externalists, that two thinkers who share all their intrinsic properties will share their thoughts. It is in this sense that externalists think that thoughts and other intentional states are not ‘in the head’: the nature of a thought is not wholly determined by what is in the head, since two thinkers could be ‘head-wise’ identical but differ in their thoughts. One main source of externalism is Hilary Putnam’s Twin Earth argument. After giving a careful exposition of the argument, Gavran raises some objections to my own criticism of it; then she argues that my alternative suggestions amount to a form of externalism. Naturally, I disagree with her on both points.

Some of the disagreement between Gavran and me is purely verbal. For example, she objects to my mentalistic version of Putnam’s principle that intension (or meaning) determines reference—‘content determines reference’—because according to her the externalist will say instead that ‘reference determines content’. But I was taking ‘determines’ to mean what Putnam means by it: essentially, it means ‘is sufficient for’. Gavran quotes Putnam as saying that the principle is that sameness in intension means sameness in extension.<sup>3</sup> This implies that a *difference* in extension/reference implies a *difference* in intension/meaning, but not that *sameness* in extension/reference implies *sameness* in intension/meaning. That latter view is quite implausible, since it fails to allow that two terms might refer to the same thing and yet differ in meaning. In other words, the externalist upholds the following principle:

<sup>2</sup> Special thanks to Nenad Mišćević for his characteristically generous and energetic organisation of the whole project.

<sup>3</sup> Note, in passing, how implausible this version of the principle is, for reasons having nothing to do with externalism. It implies that if the extension of ‘*x* is a rabbit’ were to change because of the death of one rabbit, the intension or meaning of ‘*x* is a rabbit’ would have to change. This superficial problem can be avoided by shifting from talk of extension to talk of reference.

(CDR) If two thoughts have the same content, then they have the same reference.

But not the implausible converse:

(RDC) If two thoughts have the same reference, then they have the same content.

Given what Putnam explicitly says he means by ‘determines’, the slogan ‘reference determines content’ ought to mean (CDR). After all, neither he, nor (almost) anyone else, believes (RDC). And nor does Gavran, as it happens, since she says immediately after her criticism of me that ‘reference determines content in the sense that if the reference had been different, so would have been the content’. But this use of the word ‘determines’ contradicts Gavran’s earlier use of the term, when she says that reference determines content on the externalist view. Once this terminological confusion is sorted out, then it is plain that there is no issue here between Gavran and me.

(One minor criticism, which is perhaps worth making since I have encountered it elsewhere in discussions of this matter: Gavran says that Putnam ‘built H<sub>2</sub>O into the content [of ‘water’] in order to achieve the correct reference’. It is true that Putnam says many things in ‘The Meaning of “Meaning”’, not all of them obviously consistent, and at some point he does say that meaning is a ‘vector’ with reference as one of its components. But this cannot be an *assumption* in his argument, otherwise the argument would be plainly question-begging. What is important in Putnam’s argument is how he argues that meaning is not in the head on the basis of the principle that meaning determines reference (or CDR for the mentalistic version). If he was assuming in addition that reference is *part* of meaning he wouldn’t need this argument, and his position would have been of much less interest: it would amount to an uninteresting stipulation about how to use the word ‘meaning’. A comparison: Frege does not count as someone who holds that meaning is not in the head simply because ‘Bedeutung’ can be translated into English as ‘meaning’!)

In chapter 4 of EM I argued that the CDR principle is not generally true, since it is not true for indexicals. Hence the externalist has to argue either that (a) it is true of indexicals, contrary to appearances (this is Evans’s view); or that (b) there is some special reason why CDR is not true of indexicals which does not affect the Twin Earth argument. Without an argument of either form, the opponent of externalism is under no pressure to accept the Twin Earth argument.

Gavran, however, believes that an externalist can agree with my rejection of the CDR principle. She says that in my example of two people (Alice and Bob) thinking the same indexical thought, ‘the externalist is free to say that there is an aspect of the content that is common to Alice and Bob, or me and my Twin. Putnam famously acknowledges this and names this internal component the “stereotype”.’

Here I believe that Gavran is led astray by the rather disorganised character of Putnam’s original discussion. Subsequent discussion by

externalists has—correctly, in my view—tended to avoid the issue of stereotypes. Of course, Putnam does say that stereotypes are a part of meaning. But the key question is: do they determine reference? Obviously not, according to Putnam’s argument. So, according to the principle that meaning/content determines reference, they cannot be meanings/contents. By the CDR principle, which Putnam accepts, meaning/content *must* determine reference. Abandon this and you abandon the whole Twin Earth argument. Yet it is surely very implausible to think that stereotypes are not contents or meanings! Hence the CDR principle must be false. What is shared between Twins—what Putnam calls a ‘stereotype’ but what would be called today a ‘narrow content’—does not by itself determine reference. But what is shared between Twins is surely something mental (or so I say) and the way in which they differ is something non-mental. So we should agree that what is mental is ‘in the head’ (i.e. supervenes on local properties) and does not determine reference by itself, but only in a context of thought. The difference between the Twins is not in something mental (not in the content of their thoughts) but in something extramental (whether their environment contains H<sub>2</sub>O or XYZ). This view is internalism, since it accepts the supervenience of all intentional mental qualities on local intrinsic properties. It is the essence of the internalist view defended by Fodor (1987).

I suspect Gavran is unwilling to recognise just how radical the alternative view, externalism, is. Accepting that there is some mental content in common between Twins—‘narrow content’—is accepting something which most externalists would deny. Externalists say that content is ‘broad’ or ‘wide’: in other words, it is not shared by Putnamian Twins. But perhaps Gavran’s view is really the ‘dual component’ view which says that some content is broad and some is narrow? The problem here is saying what the view really is. It cannot just be the view that the Twins are type-distinct in some ways and are type-identical in others. For everyone—internalists and externalists—agrees with this! The idea must be, rather, that their thoughts are type-distinct in some *mental* way, and are type-identical in some *mental* way. For example, it might be said that their thoughts are distinct in their context-independent truth-conditions and identical in their mental causal role. Since I do not see what point there is in calling the truth-conditions of a thought part of its ‘mental or psychological character’, I will not discuss this view further. Nonetheless I do believe that this is the direction in which Gavran is heading.

Finally, let me make a few remarks about what Gavran says about my revised version of the CDR principle:

(CDR\*) Content determines reference relative to a context.

Gavran asks why this principle is not ‘radically different from externalism’. The important point, however, is not that CDR\* is in itself externalist or internalist, but that it *cannot be used to derive an externalist conclusion from a Twin Earth argument*. Since this is a very important point, I would like to spell it out again. If one holds CDR\*, then one can say that

the content of thought supervenes on local properties—and therefore be an internalist, by definition—because then one can say that the contents of the Twins’ thoughts are *the same* on Earth and Twin Earth. In one context—Earth—the content of a watery-thought refers to H<sub>2</sub>O, and in another—Twin Earth—the *very same* content refers to XYZ. So the modified principle CDR\* allows an internalist to say that Twins’ thoughts are the same even though they are referring to different things, thus preserving local supervenience.

I do not hold, then, that CDR\* entails that ‘content is relative to a context’, as Gavran claims. For not only does this not follow from CDR\*, but it would also mean that I think the content is different in different contexts. But I think the content is the same. It is *reference* which is relative to a context, not *content*. However, Gavran’s discussion does bring out something unsatisfactory about what I said about indexicals in EM. For I did say that ‘some aspect of the content’ is different in the cases of Alice and Bob. What I had in mind was the context-independent truth-conditions—what followers of Perry and Kaplan call the ‘proposition’ expressed by an utterance of an indexical sentence like ‘I am hot’. I did not want to take sides in the debate about whether indexical sentences have two kind of content. However, one thing I have learned from Gavran’s paper is that perhaps a real internalist ought to take sides in this dispute too. If so, I should not have said that an ‘aspect of content’ is different in the case of Alice and Bob. I admit to being undecided as to how far internalism should be pushed here.

### *Reply to Dunja Jutronic*

Dunja Jutronic takes me to task over my treatment of the knowledge argument. Unlike many philosophers, I think the knowledge argument is a sound argument. But it is not a sound argument against physicalism; rather it is an argument for an interesting conclusion about knowledge: that not all our knowledge is available from a purely ‘objective’ perspective; or in other words, that there are truths which can only be known when one has had certain experiences. However, I do not think that this refutes physicalism, since I don’t think that physicalism is committed to the view that all knowledge is available from a purely ‘objective’ perspective. So when she says that I ‘argue against physicalism’ using this argument, Jutronic actually misrepresents the aim of section 28 of EM.

However, independently of this, Jutronic also finds fault with my argument that Mary in Jackson’s story learns a new proposition, expressed by ‘red looks like this’. Since her objections to this argument are independent of whether or not the argument attacks physicalism, I will consider them in more detail.

Jutronic focuses on the Lewis/Nemirow ‘ability hypothesis’ as a response to the argument. She criticises first my claim that there is not a clear distinction between ability knowledge and propositional knowledge. She agrees with Schiffer (2002), as against Stanley and Williamson (2001) and Snowdon (2003) that there are cases of know-how which are not

cases of propositional knowledge. Now this is an important issue, and I admit that I do not say enough in chapter 3 of EM to persuade someone that the know-how/know that distinction is not clearly formulated (let alone to go as far with those who say that every case of knowing how is a case of knowing that). But we can put this question to one side for the time being. For although one could use the Snowdon–Stanley–Williamson (SSW) thesis to argue against the ability hypothesis, it is not necessary to do so, as I pointed out in EM. In fact, it now seems to me a very strong dialectical requirement to place on the critic of the ability hypothesis that they have to show that the SSW thesis is true. But it turns out that one can attack the ability hypothesis in a more direct way, without assuming the truth of the SSW hypothesis.

The core of the ability hypothesis is the claim that Mary learns no new propositional knowledge. I can agree with Jutronic that Mary does learn some new abilities when she sees red for the first time. The issue which divides us is over whether she learns any new proposition. I say she does, while Jutronic says she doesn’t. However, I do not think (as she implies I do) that this proposition concerns ‘inexpressible qualities of experience’.<sup>4</sup> The qualities which one’s experience has *can* be expressed: by saying things like ‘this looks red’. Of course, these propositions cannot be understood by someone who has never seen red, but that is only to say that there are certain things you do not know unless you have had certain experiences—a view which I share with the defender of the ability hypothesis.

So what will settle the issue between Jutronic and me? She argues that I miss the point of the knowledge argument in two of the ways in which Lewis indicated in his classic paper, ‘What Experience Teaches’. The first objection is that by saying that Mary expresses her new knowledge by saying ‘Red looks like this’ I am merely saying that Mary is saying what experience she is having, and not saying what it is like to see red. But, the objection goes, Mary already knew what experience she would have before she left the black and white room. So *knowing what experience you are having* cannot be the same as *knowing what it is like to have that experience*.

I simply deny that I miss the point here. Nothing forces me to concede that by saying ‘red looks like this’ Mary is only saying what experience she is having. Rather, she is giving expression to a piece of knowledge of the experience, knowledge which she could not have had if she had not had the experience. It is this knowledge, I claim, which is her knowledge of what it is like. It is different from knowledge of what experience she is having, since I agree that she could have known that before having the experience. Hence the objection fails.

The second Lewisian objection is that I am confusing the information conveyed in a statement with the way in which it is conveyed. ‘Red looks

<sup>4</sup> I also reject Jutronic’s description of me as a defender of ‘qualia’, given the efforts I took in chapter 3 of EM to distance myself from the qualia theory! However, qualia are not really the issue here. Someone who denies that experience involves qualia (a ‘representationalist’ about experience) still has to face up to the knowledge argument.

like this' might therefore be a way of conveying the same information which is conveyed by some complex neurophysiological description. I agree, however, that nothing in the argument shows that the proposition which Mary learns could not also pick out the same real-world state of affairs that can be learned about in another way. Since I was not using the argument to show that physicalism is false, I am not worried by this feature of the argument.

The issue, rather, is knowledge. For it does not follow from the fact that the real-world state of affairs corresponding to two distinct sentences is the same, that the sentences express the same piece of knowledge. If that were true, then the knowledge that Hesperus is Phosphorus would be the same piece of knowledge as the knowledge that Hesperus is Hesperus. But it isn't. Similarly, I claim that although what is expressed by a particular utterance of 'red looks like this' might be made true by purely a physical state of affairs, this doesn't mean that it expresses the same piece of knowledge as the knowledge expressed by a sentence describing that state of affairs in the physical language of the black and white room.

So Mary learns some new knowledge; it is essentially 'subjective' knowledge; it is the kind of knowledge that you cannot have just by learning the whole of physics; but this does not mean that physicalism is false. Jutronic prefers the ability hypothesis, and finds my 'propositionalist' response implausible. I, on the other hand, find the ability hypothesis implausible, and have been given no reason to find my propositionalist response implausible. So we have a stand-off.

### *Reply to Olga Markič*

Olga Markič gives a clear account of the mental causation problem as it arises for non-reductive physicalism. She raises two questions for my account of the problem in chapter 2 of EM, and asks me to say more about the rejection of the 'completeness of physics' which I tentatively propose at the end of this chapter. I doubt whether what I have to say on this third matter will satisfy her. But let me address her questions first.

Markič's first question is about how I would fit Jaegwon Kim's (1998) reductionist solution to the mental causation problem into my classification of types of reductive and non-reductive strategies. I distinguish between ontological reduction (the identity theory) and explanatory reduction (various kinds of relationships between theories). I admit that there is much more to reduction than this; however, I do not think that Kim's latest approach falls outside this broad classification.

In his most recent work in this area, Kim has favoured a reductive account of the mind-body relation, for the reasons that Markič clearly explains. His reductive account essentially appeals to a 'functionalisation' of the mental: mental states can be characterised in functional or causal terms, that is, in terms of their typical causes and effects within a network of such states. This characterisation gives the *causal role* of a mental state. The exclusion argument tells us, in effect, that the causal role of a mental state cannot be discharged by a purely mental state. Because of

the causal closure of the physical it must be discharged instead by a physical state. In other words, we seem to have a general argument (from the exclusion argument and the functionalisation of the mental) for the identification of mental states with physical states. Hence Kim's view is a version of the identity theory and therefore a form of ontological reduction in my sense.

Although Kim does not say it explicitly, his argument for ontological reduction is very close to David Lewis's (1972) argument. Lewis argues, as Kim does, from a functional analysis of mental concepts to an identity theory of mental properties. The functionalism of Lewis and Kim, therefore, is not in opposition to the identity theory; rather, it is an assumption in the argument for the identity theory. There is another form of functionalism, associated with Putnam's early work, according to which mental properties are 'functional properties' and are not reducible to physical properties. Though this is often how 'functionalism' is understood, it is plainly not how Kim understands it, because he rejects the idea that the functional/non-functional distinction is a distinction between properties; rather, it is a distinction between predicates. This is another reason, in fact, why Kim and Lewis are so similar.

Markič's second question to me is about the distinction between emergence and non-reductive physicalism. Referring to an earlier paper of mine, 'The Significance of Emergence' (published in 2001 but actually completed in 1997) she points out that there I drew the distinction in epistemological terms, but that it is more usual for the distinction to be drawn in terms of whether a theory accepts 'downwards causation'. In chapter 2 of EM, I claim that an emergentist is someone who accepts the merely nomological supervenience of the mental upon the physical, as opposed to a physicalist who accepts the metaphysically necessary supervenience. It appears then that lots of things have been said about what emergence is—not all of them obviously related—and I am grateful to Markič for giving me this opportunity to sort a few things out.

The issue is how to distinguish between the kind of theory which denies the identity theory yet accepts physicalism (non-reductive physicalism) and the view which deserves the name of emergence. In my 2001 paper I was sceptical about whether there is a real metaphysical distinction here; but I now think that this scepticism is wrong. Part of what I said there was right: a physicalist must believe that all facts must in some way be explicable in physical terms. The emergentist denies this. I would now express this by saying that a physicalist must hold that there is an explanatory reduction of the mental. Hence all forms of physicalism are reductive in one way or another: either ontologically or explanatorily.

But there is more to it than this. For physicalists, if they want to answer the mental causation problem, must hold that the physical facts fix all the facts. This need not be because all facts are identical with the physical facts (the anti-ontological-reductionist physicalist denies this); it is rather because of the weaker idea that whatever other facts they are, they must be 'nothing over and above the physical facts'. Hence all God

has to do to create the facts is to create the physical facts. The rest come for free. To my mind, such a view (whether ontologically reductionist or not) deserves the name of physicalism. An emergentist, then, is someone who holds that all particulars are physical, but denies that the physical facts fix all the facts in this sense. The extra facts, although emergent, do not ‘come for free’, they are something ‘over and above’ the physical facts, even if they are not facts about some non-physical substance. I admit that these ways of describing the view are rather imprecise, so I prefer the following: the non-reductive physicalist holds that the mental facts supervene on the physical facts with metaphysical necessity, whereas the emergentist holds that they supervene only nomologically. This conforms to the terminology of current debate: the emergentist agrees with someone like Chalmers (1996) that facts about consciousness do not supervene on the physical. Chalmers’s argument is an argument against the view that there is a necessary determination relation between the mental and the physical: this is physicalism as I am construing it.

A physicalist, then, must hold either an ontological reduction (like Armstrong or Davidson) or metaphysically necessary supervenience with an explanatory reduction (like Jackson or Block). My term ‘explanatory reduction’, as Markič says, is another way of talking about what Terence Horgan means by ‘superdupervenience’. An emergentist denies ontological reduction and also denies the metaphysically necessary supervenience and the inevitability of an explanatory reduction. (These are only necessary conditions for being an emergentist of course; many other views will agree with these anti-physicalist thoughts.) Hence the situation is more complex than I said in my 2001 paper, and I am grateful for Markič for giving me the opportunity to set the record straight.

Markič asks me, finally, to say more to justify my scepticism about the completeness (or causal closure) of the physical. This is a very difficult debate, which I enter with some trepidation. Part of the problem lies with the definition of the causal closure principle. Papineau is well-known for advancing one very clear version: that physical causes suffice to fix (or fix the chance of) all effects. So let’s begin with this version of the principle as our example. At the end of chapter 2 of EM, I tried to make this version of the principle consistent with the reality of mental causation by spelling out in a bit more detail what ‘sufficient’ must mean. (Let’s ignore indeterminism for simplicity.) I proposed that no causes are absolutely sufficient for their effects; causes are only ever sufficient in the circumstances. But if we reject, as we should, the distinction between cause and circumstance, then the circumstances will include other causes too. So, I said, a cause can be sufficient for an effect only given all the other causes which are present. I then proposed that just as a mental cause can be sufficient for an effect, given the other (physical) causes of it, so a physical cause can be sufficient for an effect, given the other (mental) causes of it. The completeness of physics is then the claim that every physical effect has a physical cause which is sufficient in *this sense*: i.e. in a sense which also allows mental causes to be active in bringing about the

same effect, though not by overdetermining it (remember that the mental and the physical are linked by law). In this sense, I claimed, an emergentist could believe in an innocuous version of the completeness of physics.

However, I doubt whether this attempt at conciliation would satisfy Markič and other physicalists. The reason—and this is mere conjecture—is that they think that the physical causes are sufficient in a different ‘unconditioned’ sense: not just sufficient in the circumstances but sufficient *tout court*. If this is their response, then all I can say is that I don’t believe that any causes (except perhaps Mill’s ‘entire state of the universe’ before an effect) are sufficient *tout court*.

There is one final point I would like to make about the completeness of physics. I believe that, faced with a choice between two hypotheses, it is better to believe the weaker (or less committing) hypothesis if one has insufficient evidence for the stronger. Emergentism, as I present it in EM, is a weaker view than physicalism. Whether we have reason to believe the stronger view in this case depends on two things: the reasons we have to believe in the metaphysically necessary determination of the mental, and whether we have reason to believe in the unconditioned version of the completeness of physics. As for the former, the jury is still out: much of the current debate about consciousness turns on this question. Nonetheless, there are good arguments provided by functionalists like Kim and Lewis about where one might start in defending the necessary determination thesis.

When it comes to the non-innocuous version of the completeness of physics, though, it seems to me that the burden of argument is still on those who believe this strong thesis. As far as I understand the matter, the thesis is not part of what is known in current science, and it involves a huge empirical conjecture which few of us would have much idea about how to establish. I have not yet been given a good reason to believe in the completeness of physics, despite the attempts of Papineau and others, and therefore I would prefer to do without it if I can. This is not an ‘anti-scientific’ attitude; it is just epistemic caution in the face of what I take to be the scientific and non-scientific evidence.

### *Reply to Nenad Mišćević*

Nenad Mišćević’s interesting paper is a defence of a response-dependent or dispositionalist account of colour, but one which appeals to an intentionalist account of colour experiences (the ‘responses’ in terms of which colour is defined). Mišćević’s view, as he says, is entirely consistent with the intentionalism about perception I defend in chapter 5 of EM. Of course, an account of colour is one thing, an account of perceptual experience is another. In my book, I only offered the beginnings of an account of experience (including some remarks about the experience of colour). But naturally one would expect the accounts to have something to do with one another. I do not have any worked-out view on colour, so I am grateful to Mišćević for offering the response-dependence-intentionalist (RI) account.

Nonetheless, I cannot bring myself to accept it, despite its obvious merits.

An initial (minor) difficulty with some of the things Mišćević says about RI is that the thesis is insufficiently general. His official description of the RI thesis says that ‘being red ... is being such as to cause the response of visaging phenomenal red in normal observers under normal circumstances’. But in other places he talks, as others do, of redness being a property of *surfaces*. Yet light can be coloured too, as well as surfaces. So a full account of colour must explain how both light and surfaces can be such as to give rise to certain kinds of experiences in certain circumstances. This will, of course, make the categorical base of the disposition much more heterogeneous than a property which is merely confined to surfaces. I don’t say this is a problem for Mišćević’s account; I draw attention to it only because colour is often discussed in these contexts as if it were only ever a property of surfaces.

My main worry about accounts like RI, however—and the thing which led me to say (in EM: 147) that accounts like it are ‘seriously flawed’—is the dependence in the definition on the concept of normality. Mišćević asks me to expand on this claim, and I am happy to do so. In EM section 44 I discussed the traditional inverted spectrum argument, which some have used to prove the existence of qualia. Since I reject qualia, and all their works and empty promises, I have to say something about the inverted spectrum case. My proposal is that the best way to save the appearances is to insist that one of the subjects in an inverted spectrum case is misrepresenting the world in their experience. The difficulty with this is that it seems to be an inessential feature of the thought–experiment that there are very few ‘inverts’: what do we see if 50% of the population’s colour experience were inverted relative to the other 50%? This strikes me as a genuinely deep problem: on the one hand, realists about colour want to say that colour judgements are made true by how things objectively are, which colours things have; but on the other hand, the 50/50% population spectrum–inversion possibility says that these judgements can’t be straightforwardly true or false; they have to be true–for–50%. But this is patently not how we normally think about colour. If we then say in the style of Mišćević that ‘red’ means ‘being such as to cause the response of visaging phenomenal red in normal observers under normal circumstances’ then I want to know how to answer the question: what is a normal perceiver? Is it someone from this 50% or someone from the other 50%? Or is it that some things are red for some people and not red for others, because of the population they belong to? Certainly there are real cases of groups of people of whom such things are true; but I think our intuitive reaction is that such people are *misperceiving* the world in a certain obvious way. But how can one say this in the 50/50 case if one has no standard of what is normal?

Mišćević is well aware of this objection to views like RI. His response is that the question of the definition of normal raises a problem which all theories ultimately have to face. I agree that it seems to be a problem for

most of the plausible theories; and within the dialectic of arguments about colour, Mišćević’s point is effective. But this fact in itself makes me—someone with no settled view on the nature of colour—no more sympathetic to Mišćević’s RI, despite its merits. Rather it makes me withhold judgement on all (broadly) reductive theories of colour. My problem, then, is that I have no satisfactory intentionalist answer the inverted spectrum argument until I have an adequate theory of colour.

### *Reply to Puškarić*

Ksenija Puškarić raises four questions about my characterisation of intentionality: (i) whether states of mind like depression can really be said to have intentional objects; (ii) whether the ‘feel’ of a bodily sensation goes with the intentional mode or the intentional content; (iii) how the notion of *non–conceptual content* should be explained; (iv) whether in certain cases, the notions of intentional object and intentional content merge into one another. Questions (i), (ii) and (iv) are all about general features of intentionality, so they can be dealt with together; after having said something about these general features, I will make a few remarks about non–conceptual content.

Puškarić seems to think that the fact that someone can be depressed without knowing what they are depressed about is a reason for thinking that depression is not intentional. But this would be so only on the assumption of the general principle that subjects must always be able to know, or even to say, what the intentional object of their state of mind is. Yet only the slightest reflections on the limits of subjects’ self–knowledge will show this general principle to be false, for reasons entirely independent of the truth of intentionalism. Given this, then it is open for the intentionalist to say that the answer ‘nothing in particular’ could simply be a statement of the subject’s lack of self–knowledge. An alternative is to say that ‘nothing in particular’ is also a way of expressing the idea that *things in general* is what the state of mind is directed on: nothing in particular, just things in general, the whole of the subject’s world. Intentionalism allows both these descriptions of the situation and is not obliged to choose between them. The important thing is that depression is not a simple monadic quality, but has an intentional structure.

The structure of intentionality is the topic also of Puškarić’s second question. According to my account of pain in chapter 3 of EM, a pain involves an awareness of a part or region of the body, the part that hurts. The part or region is the intentional object, under a certain aspect, and the way in which one is aware is the intentional mode. The parallel is with sense–perception: when a rabbit is seen, for example, the rabbit (under a certain aspect) is the intentional object of the intentional act; and the seeing is the intentional mode. But Puškarić is dissatisfied with this kind of account of sensation: what, she asks, is the mode in the case of pain? Is it feeling or hurting?



The question is a good one. But I don't think it raises a serious problem for my view. My account is intended to dismantle the notion of a part of one's body feeling to be a certain way. The way something feels to be breaks down into two components: the aspect under which the thing consciously appears (the aspectual shape or intentional content) and the mode in which it appears (the intentional mode). The intentional mode, in the case of pain, is a certain conscious way of feeling the part of the body which hurts. But since the content of the state is also involved in the entire conscious state of hurting, it would be wrong to say that the intentional mode is the hurting. Rather, the conscious character of that which we call hurting is the result of the content and the mode. I take the parallel with outer perception very seriously therefore: the conscious character of a visual experience, for example, is the result of both what is seen (the content) and the fact that it is seen, rather than heard or touched (the mode).

Given this, it was a little misleading of me to say that the content of a pain in my leg might be given by the sentence 'my leg hurts', since that makes it appear that the 'hurting' side of things, so to speak, is in the content rather than the mode. But, as Puškarić's question brings out, the truth is rather that the entire state of mind (mode plus content) is *expressed* by such a sentence. Similarly, if I say 'I see a herd of kangaroos' this can be the *expression* of an entire visual state, and that does not mean that the concept of seeing enters into the content of the state. I should not have said, then, that the content of a state of mind is what one would put into words, if one has words in which to put it. I should have said that one's words can express the content of a state of mind; but in addition they can express the mode too.

Puškarić's other question in this area is about how to distinguish content and object. She asks:

Let's say that I open my eyes and see nothing but darkness. Then my mind is directed upon the darkness. The darkness is the intentional object of my perception because it answers to the question 'what is my perception directed on?'. On the other hand, the darkness is the content of my perception because it also is the answer to the question 'what is in my mind?'. Here, the content and the object seem to be the same. The question is how can we distinguish the content from the object?

It is true that sometimes the distinction between object and content can seem rather hard to grasp, given the fact that whenever one has to say what the object of a state of mind is, one always has to describe it in some way, and therefore give the object under some aspect. If I say that the object of my thought is the Duke of Wellington, then I have also given in some sense the aspect under which I was thinking about him. Notice too that 'the Duke of Wellington' could be an answer to the question about what is in my mind—the content of my thought—as Puškarić says. But one way I distinguished between content and object, in chapter 1 of EM, is by saying that the same object can be represented under different as-

pects. The Duke of Wellington can also be thought of as the victor of Waterloo, and these two ways of thinking of him yield different contents concerning the same object. This is enough to get an initial grip on the distinction between content and object. To apply this to Puškarić's example of the darkness: the darkness might be identical with a condition of my room in which my bed is there but invisible because of an absence of sufficient light. But it need not seem to me to *be* this condition, since I might be aware of the darkness while being unaware that I even have a bed. Hence although 'the darkness' might give the content and the object of my state of mind, these are not identical if the same object can be given in different ways.

Finally, Puškarić raises some questions about the notion of non-conceptual content. In chapter 5 of EM I claimed that we should understand the phenomenal character of phenomenal experience partly in terms of the fact that its content is non-conceptual. Like many who have encountered this debate, Puškarić is suspicious of the very idea: 'it is not clear what non-conceptual content is at all, we only know what non-conceptual content is not. By giving a kind of negative definition, the conditions for making a content non-conceptual one, remain unknown'. I disagree. If we have some idea of what a concept is, then we can say what it is for a state to be conceptual, according to my definition. For we can say that a state is non-conceptual when (a) it is intentional or representational, and (b) not conceptual, by the earlier definition. This makes the conditions for being non-conceptual perfectly explicit, so long as we know what makes something intentional and what makes something conceptual.

Someone might doubt, with Puškarić, whether the issue of what makes something conceptual has been made sufficiently clear. Puškarić herself asks about the relation between having a concept of something and (a) having the capacity to recognise it; and (b) having a word for that thing. I myself doubt whether either or both of these ideas can exhaust what it is to have a concept. If having a concept of X is just a matter of having a word for X in one's language, then everyone (*pace* John McDowell) will agree that perceptual experiences have non-conceptual content according to my definition, since almost everyone will agree that we can perceptually experience things for which we have no words. So if there is a debate about whether experiences have non-conceptual content, which there seems to be, then having a concept cannot just be a matter of having a word in one's language. Similarly, I don't think that having a concept of X is simply a matter of mentally representing X. If it were, then all content would be (trivially) conceptual, since content simply is what is represented by a mental representation. But the question of whether the content of experience is non-conceptual cannot be answered so easily in this way either. The relevant notion of a concept must lie somewhere in between the demanding sense—a word in one's language—and the trivial sense—mental representation. The notion of a recognitional capacity certainly *would* occupy such a middle position, and it would be an interesting pro-

posal that all concepts are recognitional capacities. But the proposal is, I think, mistaken; since understanding 'recognition' in the normal way, the proposal can only be true of the kinds of concepts which are derived from perceptual experience itself. And unless there is some reason to hold the empiricist view that all concepts are derived from perceptual experience, then there is no reason to accept the proposal. For it is plain that there are many concepts which have little connection to experience, and therefore little connection to recognition in the normal sense.

A better proposal would be to locate the notion of a concept in relation to the notions of reasoning and rationality. To have a concept of X would be to be able to reason about X, which minimally requires having beliefs about X, being sensitive to what might rationally support these beliefs and what might rationally overturn them. Belief is the paradigm of a conceptual state, and beliefs are the kinds of mental states for which reasons can be given. Perceptual experiences and bodily sensations are not like this: we do not give reasons for our experiences and our sensations. It is this distinction between those states of mind to which the notions of reason and rationality applies, and those to which they don't, that I think should form the basis of the distinction between the conceptual and the non-conceptual. But it goes without saying that a lot more needs to be said in order to make this distinction precise, and to this extent the distinction between conceptual and non-conceptual content still needs to be properly articulated.