

# Realism about cognitive science

Mark Sprevak

## 1. Introduction

This chapter is about a puzzle. Realism about  $X$  is often glossed as the claim that  $X$ s are mind independent:  $X$ s exist, and have their nature, independent of human beliefs, interests, attitudes, and other mental states.  $X$ s are out there, getting on with it, independent of human minds. How then should one understand realism about the mind? Having an answer to this is important if one wants to be a realist about cognitive science. The subject matter of cognitive science includes mental states, mental processes, and mental capacities. None of these are mind independent. To say otherwise would be an obvious error. But then how can one be a realist about them? This is our puzzle. My solution will be to distinguish between two types of mind dependence that pertain to cognitive science. One type is trivial and follows from the nature of the subject matter. The other type is non-trivial and it is the true point of contention between a realist and an anti-realist about cognitive science. My aim in this chapter is to identify that point of contention.

In Section 2, I describe different kinds of realism that one might adopt in cognitive science. In Section 3, I argue that the kind of realism that asserts mind independence has a special role to play in cognitive science. In Section 3, I present the puzzle about this kind of realism: it seems impossible to be a realist of this kind about cognitive science. In Section 4, I examine three solutions to the puzzle. Each draws the distinction between a trivial and a non-trivial kind of mind dependence in a different way. My favoured proposal derives from the observation that theories in cognitive science aim to explain mental phenomena by associating them with structured complexes – for example, computations, mechanisms, networks, or causal chains. I claim that realism in cognitive science should be understood as a claim about the *individuals and relations* that compose those structures not about the *entire complexes taken as whole*. Mind dependence about the wholes (hypothesised to realise, constitute, or otherwise compose mental processes) is trivial. Mind dependence about the parts and relations that make up those wholes is not. This is the true point of disagreement between realists and anti-realists about cognitive science.

## 2. Kinds of realism

Realism is not a single claim but a range of possible claims that could be made about a range of subject matters. One might be a realist about one type of entity or subject matter and

an anti-realist about another. One might be a realist about electrons but an anti-realist about beauty marks. ‘Local’ kinds of realism should also be distinguished from ‘global’ versions. A global realism asserts realism about all or most subject matters of the mature sciences. I will not consider global realism here. My concern is with specific realist claims made about entities in cognitive science.

Within a local domain, a realist may make a range of different claims. Realist/anti-realist disputes take on a different character depending on which claim is at issue. In this section, I highlight six possible types of realist claim: claims regarding existence of an entity, the nature of that entity, referential semantics for the discourse that purports to talk about that entity, truth or approximate truth of that discourse, evidence for truth of that discourse, and mind independence of the entity. These claims are distinct and a realist may assert or deny them in various combinations.

First, *existence*. On this view, realism commits one merely to the existence of the relevant entity. Fodor (1975) is a realist in this sense about beliefs. The relevant kind of anti-realism would be eliminativism. Churchland (1981) holds this position about beliefs.

Second, *nature of the entities*. Assuming that *Xs* exist, what sort of things are *Xs*? One idea is that *Xs* are discrete individuals. Fodor (1975) is a realist in this sense about beliefs. Beliefs are countable particulars that occur and may re-occur inside someone’s head. The relevant kind of anti-realism would be a deflationary view of the relevant entity. Dennett (1991b) argues that beliefs are not individuals but shifting and transitory patterns in and around agents that observers may exploit for predictive or explanatory gain.

Third, *referential semantics*. If one is a realist about *Xs*, then the relevant part of the discourse that purports to talk about *Xs* should be understood as having a referential semantics. Fodor (1975) is also a realist in this sense about beliefs. If we say, ‘Abby has the belief that beer is in the fridge’, we refer to some thing that Abby has. According to Fodor, this is a sentence in the language of thought tokened inside her head. The relevant form of anti-realism would be a non-referential semantics for the relevant discourse. Ryle (1949) advocates this kind of anti-realism about beliefs. When we say, ‘Abby has the belief that beer is in the fridge’, we do not refer to any thing (or even a deflated thing) that Abby has. Instead, we merely convey to our listener a warrant to make inferences about, among other things, Abby’s behaviour.

Fourth, *truth*. If one is a realist about *Xs* then the relevant part of the discourse aims to tell the truth. Block (2007) advocates this form of realism about phenomenal consciousness. Experiments to study phenomenal consciousness involve reports from human subjects about the occurrence or non-occurrence of their subjective phenomenal experiences (reporting, for example, that they experience red). We should, according to Block, understand these reports

as aiming to tell the truth about those episodes of conscious experience. In contrast, Dennett (1991a) argues that we should be fictionalists about phenomenal consciousness. Reports of experiencing red should be understood, not as aiming to tell the truth about the occurrence or non-occurrence of conscious episodes, but as a roundabout way for the subject to express to herself and others that her cognitive system has detected a highly disjunctive physical property (such as redness). An anti-realist denies the truth-seeking character of discourse about *Xs* but may maintain that talk about *Xs* has other virtues.

Fifth, *evidence*. If one is a realist about *Xs*, then we have justification for the truth (or approximate truth) of the relevant part of the discourse. Block (2007) is a realist in this sense about phenomenal consciousness. Subjects' reports of conscious experience not only aim to tell the truth, we also (normally) have justification that they *are* true. Significantly, the justification holds even under unusual presentation conditions such when visual stimuli are flashed very briefly to subjects in Sperling (1960)'s experiments (a grid with characters is briefly presented followed by a visual mask). The relevant form of anti-realism would involve epistemic caution about the relevant claims. Irvine (2012) claims that we lack justification for believing the reports of subjects about their experiences in the context of Sperling's experiments.

Finally, *mind independence*. This realist claim is different to the others. Like the second claim, it concerns the nature of the entities. But it is about a different property. The question here is not about their nature as individuals but about their degree of mind dependence: does that entity depend, for its existence or nature, on the mind? To see why this might matter, consider that all our knowledge of the world is mediated by our minds. We cannot see the world untainted by human conceptual, motivational, and other cognitive systems. We may attempt to counteract the effects of our cognitive makeup by taking into account its hypothesised distortions and biases. But seeing the world 'as it is', without the interpolation of the human mind, is impossible. This invites a question. Which parts of our knowledge correspond to entities and properties that are *really out there* and which are *constructions of, imposed by, or artefacts of*, our minds? Some entities seem to exist and have the properties that we attribute to them independently of the way we conceive of them. Fundamental particles in physics like electrons may be like this. If our minds were not to exist, or if they were to have a radically different makeup, electrons would (likely) continue and have unchanged properties. Other entities appear to be partly constructions of our minds. Beauty marks may be like this. Whether a specific skin colouration counts as a beauty mark is not a mind-independent matter. It depends on how that colouration strikes, or would strike, our mind and fit with our visual preferences – whether that patch *looks beautiful to us*, an agent with our specific cognitive and motivational makeup. If human minds were not to exist,

or if they were to have a radically different cognitive makeup, the distribution of beauty marks in the world would be different.

Realism seems *prima facie* plausible for electrons and anti-realism seems *prima facie* plausible for beauty marks. Other cases are less clear. Are *biological species* mind independent? Or do they merely reflect biases of our mind to group individual organisms in certain ways? Are *biological functions* mind independent? Or are they merely properties of organisms that happen to strike our minds as particularly salient or interesting? There seems no reason why the realism/anti-realism question cannot be asked about entities in cognitive science. Among those entities are neural computations. Neural computations are used to explain human mental processes and mental capacities. Specific mental processes – for example, specific kinds of decision making – are explained by saying that the brain of the person concerned performs specific neural computations. Cognitive science references neural computations to explain mental life. Should one be a realist or an anti-realist about neural computations?

Fodor (1980) is an example of a realist about neural computation. Suppose that Abby's brain performs a specific computation which realises her decision-making processes that determines on a specific occasion whether Abby goes to the fridge to get another beer. According to Fodor, whether Abby's brain performs this computation, or any computation at all, has nothing to do with *how we view* Abby. Whether Abby's brain performs this computation or not is determined entirely by facts about Abby's brain. Burge (1986) advances a similar realist claim but holds that the neural computation is determined by a broader base of mind-independent facts: not only Abby's brain but also her causal relationship to her environment determine which computation her brain performs. Despite their disagreement, Fodor and Burge agree on realism about neural computation: neural computations are not a grouping that is dependent on how we human agents view a brain – a grouping that is perspicuous to us but not reflective of any objective distinction in the world. Neural computations are 'out there'. The aim of computational cognitive science is to discover and describe those neural computations. This description may be right or wrong, but it is so independently of how human agents conceive the world.

In contrast, Putnam (1988) and Searle (1992) argue for anti-realism about neural computation. According to Putnam and Searle, neither Abby's brain nor her brain plus and her causal relationship to her environment determine whether her brain performs a specific computation. Absent consideration of *how we view* Abby's brain, there is no fact about whether Abby's brain performs one computation rather than other, or any computation at all. Neural computations are 'observer-relative': they are constructions based on what strikes us as a perspicuous grouping. If human minds were not to exist, or if they were to have a

radically different makeup, the distribution of neural computations would be different. Neural computations are like beauty marks and not like electrons: they are not ‘out there’ waiting to be discovered. They are a construction that reflects the specific way in which humans conceive the world.

You might have already spotted a glimpse of our puzzle from this description. For electrons and beauty marks, the question about mind dependence can be posed in a relatively straightforward way: either the entity depends on the human mind or it does not. These two options, whatever else one might say about them, seem both exhaustive and coherent possibilities. Not so for neural computations. Neural computations are intimately connected to mental life. According to reductive theories in cognitive science, neural computations realise or otherwise constitute mental life. This makes realism about neural computations hard to state. If neural computations realise mental life, how can those neural computations *also* be mind independent? Consider the disagreement above. Fodor, Burge and other realists about neural computation cannot be saying that Abby’s neural computations are mind independent *simpliciter*. The neural computations in question – the neural computations that underlie Abby’s decision making about the beer – clearly depend on at least one mind: Abby’s own. If Abby’s mind were not to exist, or to have a radically different nature, those neural computations would differ. Similarly, Putnam and Searle cannot be saying that Abby’s neural computations are *in some way or other* mind dependent. For that would be trivially true. Her neural computations cannot exist, or have their nature, independently of her mental life. Both a realist and an anti-realist must agree that Abby’s neural computations are dependent *in some way* on her mind. The realist/anti-realist dispute about neural computation therefore cannot be about mind dependence *simpliciter*. Something else must be going on. Identifying what this is – what is at stake in the realist/anti-realist dispute in cognitive science – is our puzzle.

Earlier in this section we saw that a realist about cognitive science need not endorse mind independence. We saw five other ways to be a realist. This suggests a quick rejoinder to the puzzle. If mind independence is problematic, why not simply refrain from asserting it and instead pursue some other form of realism? In the next section, I argue that, while there is nothing wrong with the five other forms of realism, this strategy would have a significant cost. Cognitive science needs a mind-independence claim to fulfil one of its wider ambitions: namely, to naturalise the mind.

### 3. Why care about mind independence?

The world contains at least two kinds of phenomenon: mental phenomena – involving beliefs, sensations, ideas, concepts, thought processes, judgements, and so on; and physical phenomena – involving bodies, brains, atoms, molecules, cells, and so on. The two appear to be related: changes in one correlate with changes in the other. But the exact nature of the relationship is unclear. In particular, it is unclear whether mental phenomena are *sui generis* entities or whether they somehow ‘arise from’ the physical. Mental phenomena are puzzling not because they are complex – although they are complex – but because we do not know how they relate to the physical world.

Some theories in cognitive science aim to bridge this gap. Those theories pair a specific mental phenomenon with a non-mental phenomenon. The non-mental phenomena often have special properties: the states *perform computations, represent, process information, carry error signals*, and so on. Specific instances of decision making, for example, are paired with specific neural computations (Schultz, Dayan, and Read Montague 1997; Gold and Shadlen 2001; Gold and Shadlen 2007; Rangel, Camerer, and Montague 2008).

Such theories propose a relationship between the mental and the non-mental that goes beyond that of mere correlation. The precise details differ between cases, but two general observations can be made. First, the association between the mental and non-mental has a *non-trivial modal extent*. The mental and non-mental are reliably correlated: they co-occur reliably across a wide range of circumstances including conditions not experimentally tested. Precisely how far the modal dimension extends – across every possible world, across worlds with the same physical laws, across worlds with the same natural laws – is open to question, but we can be sure that the association has a non-trivial modal extent. The second observation is that the non-mental member of the relationship would *produce the same effects as the mental member*. Whichever scientifically-relevant properties are associated with the mental phenomenon – its characteristic scientific ‘effects’ – those properties would also be produced by the non-mental phenomenon. The non-mental member could substitute for its mental counterpart without change in scientifically-relevant properties. For example, the scientifically-relevant properties associated with decision making include patterns in behaviour, patterns in error making, how uncertain evidence is weighed, reaction times, and characteristic downstream neural effects. A potential non-mental partner would not only need to co-occur with specific instances of decision making but also to produce those characteristic properties. The drift-diffusion model, for example, aims to provide not just a neural correlate of decision making but also show that this correlate would produce the characteristic properties associated with decision making regarding reaction times, weighting of evidence, and susceptibility to errors (Gold and Shadlen 2007).

If a non-mental phenomenon co-occurs with a mental phenomenon across a wide range of circumstances and it also generates all the scientifically-relevant properties associated with that mental phenomenon, then we are in a position to advance a reductive claim. Indeed, in light of considerations of parsimony (and pending reasons otherwise), it seems rational to do so. Rather than hold that the mental phenomenon and non-mental phenomenon are two distinct entities that happen to co-occur, we may reduce one to the other. One might hypothesise that the mental and non-mental bear some reductive relation – perhaps *identity*, *realisation*, *constitution*, *grounding*, or another – to each other. For example, one might claim that decision making *is* a specific neural computation, or that decision making is *realised* by a neural computation, or that decision making is *grounded* by that neural computation.

The theories in question identify a reductive base for a mental phenomenon. The details of this reduction may be left open (*identity* vs *realisation* vs *constitution* and so on). But the general idea of finding some non-mental base that would produce the scientifically-salient effects of the mental phenomenon is shared. One pairs a mental phenomenon with a non-mental phenomenon in such a way that the non-mental phenomenon is sufficient for, or somehow produces, the scientifically-relevant properties of the mental phenomenon.

Successful reductions appear to be the golden road to naturalising the mind. By ‘naturalising the mind’ I mean explaining scientifically-relevant properties of mental phenomena in terms that do not refer to, or otherwise already presuppose, mental life. A naturalising explanation is one that takes as its *explanandum* some scientifically-relevant property of a mental phenomenon (for example, some property of decision making) and gives as its *explanans* an account that does not refer to, or otherwise already presuppose, mental life (for example, an *explanans* exclusively in terms of neural computations, physical inputs, and physical outputs). Naturalising requires realism about the subject matter of the *explanans*. One needs to be a realist – in the sense of defending mind independence – about the entities cited by the *explanans*. Consider the alternative. Suppose for the sake of argument that anti-realism about neural computation is correct. Explaining decision making by appeal to neural computation would in such a case not serve to naturalise that phenomenon. The reason is that this form of explanation would not explain the phenomenon in non-mental terms. It would explain the phenomenon in terms of entities that depend (albeit in perhaps not an immediately obvious way) on minds for their existence and nature. Neural computations, according to the anti-realist, are like beauty marks: partial constructions from human beliefs, interests, attitudes, and other aspects of our mental makeup. Explanation of mental phenomena in terms of neural computation would not be an explanation that does not refer to, or already presuppose, mental phenomena. It would not, in the sense of ‘naturalising’ above,

be a naturalising explanation. One might of course still explain decision making in terms of neural computation. But one should not mistake this for a naturalising explanation: one has not shown how decision making arises from non-mental ingredients. Rather, one would have explained by moving in a non-reductive circle: explain a mental phenomenon in terms, *inter alia*, of other mental phenomena. Nothing wrong with this – traditional forms of psychology have been doing this for a long time. But it does not fulfil cognitive science’s ambition to naturalise the mind. In order for the naturalising strategy pursued by cognitive science to work we need to be realists – and specifically, realists who defend mind independence – about the subject matter of the *explanans*.

Realism about the subject matter of cognitive science is not an idle intellectual kink. Realism is needed for explanations within cognitive science to serve the project of naturalising the mind. I repeat that naturalising is not the only goal of cognitive science. It is perfectly coherent to pursue cognitive science without any naturalistic ambition. But giving up that ambition in the face of the puzzle is not an action to be taken lightly. Consider what pay-off we would miss out on: finally understanding how the mind arises from non-physical ingredients. Rather than abandon this form of realism, let us instead examine the problem more closely and then see how to solve it.

#### 4. The puzzle about mind independence

Reductive theories in cognitive science aim to pair mental phenomena with non-mental phenomena. A reduction of this kind would appear to open the door to naturalise the mind. However, this can only work if one can be a realist – in the sense of defending mind independence – about one side (the ‘non-mental’ side) of the relation. The problem is that the preceding two claims – (i) mental phenomena reduce to non-mental phenomena, and (ii) the non-mental side of the relation is mind independent – are incompatible. This, in a nutshell, is the problem. Let us examine it in more detail.

First, consider what happens if the reduction in question is *identity*. Assume, for the sake of argument, that some instance of human decision making *is* a specific neural computation. To naturalise decision making, we would need to be realists about this neural computation: we would need to assert that it is, in an appropriate sense, mind independent. But how could that neural computation be mind independent? If decision making *is* a neural computation, that neural computation must *ipso facto* be mind dependent. Being identical to a mental phenomenon entails mind dependence if nothing else does. If human minds were not to exist, or if they were to have a different makeup, the existence and nature of this neural



computation would be different. What better reason could there be for thinking that something is mind dependent than it being *identical to* a mental phenomenon? But if the neural computation is mind dependent, then anti-realism seems to be true of it and realism seems to be false. The reduction – decision making is a specific neural computation – seems to preclude realism about the reductive base.

What if the reductive relation in question were *realisation*? Identity is symmetric: if  $X$  is identical to  $Y$ ,  $Y$  is identical to  $X$ . Perhaps it is the symmetric nature of identity that is the source of the problem. The realisation relation is asymmetric: if  $X$  realises  $Y$ ,  $Y$  does not realise  $X$ . Perhaps an asymmetric relation (like realisation) would allow us to confine the mind dependence to one side of the relation: ideally we would like the mental phenomenon to depend on its reductive base but the reductive base *not* to depend on the mental phenomenon. Unfortunately, this hope is misplaced: despite realisation being an asymmetric relation the reductive base cannot occur *independently* of its mental phenomenon, which is what the realist requires. Suppose that an instance of decision making is *realised by* some neural computation. If it is realised by that neural computation, then that neural computation is sufficient for that instance of decision making to occur.<sup>1</sup> The occurrence of the neural computation is sufficient to bring about an instance of decision making and all associated scientifically-relevant properties; otherwise, it would be unclear that we had identified a reductive base for decision making at all. *If* this neural computation were to occur, *then* the relevant decision-making process and its associated properties would occur too. Moreover, this conditional holds over a non-trivial range of modal scenarios (the precise extent of which would be determined by the realisation relation in question). Therefore, the neural computation is tied to the mental process in a modally rich way: the computation cannot occur without the relevant decision making occurring (although of course realisation would allow for the decision-making process to occur in other ways too). But then the neural computation cannot be mind independent. The neural computation cannot occur independently of how matters stand in the mental realm. In particular, it cannot occur *without* the associated mental process also occurring. The reductive base may not be realised by the mental process but the reductive base is nevertheless not mind independent. Let us see the point schematically. Suppose that a reductive base,  $B$ , realises some mental process,  $M$ .  $B$  is tied in the modally rich way prescribed by the realisation relation to  $M$ .  $B$  cannot occur (over some non-trivial range of modal circumstances) without  $M$  also occurring. But this means that  $B$  is not mind independent.  $B$  cannot occur independently of mental phenomena. If

---

<sup>1</sup> I assume we are considering the total realiser here (Shoemaker 2007). Changing to talk about the core realiser would not improve matters as core realisers have further worries pertaining to their mind dependence (Wilson 2001).

human minds were not to exist, or if they were to have a radically different makeup, then the facts about *B* would be different. The realist's claim that *B* is mind independent is simply incompatible with the claim that *B* realises *M*.

Other reductive relations – such as grounding or constitution – suffer from the same problem. For any reductive relation, the reductive base should, in some modally nuanced sense, be sufficient for the mental phenomenon. It should be sufficient to ‘bring about’ that mental phenomenon. The specific sense of ‘bringing about’ will be cashed out in different ways by different relations (identity, realisation, grounding, constitution, and so on). But irrespective of these differences the reductive base will be *sufficient for* the mental phenomenon – otherwise, why think we have found a reductive base at all? If it is not sufficient for the mental phenomenon, then we would merely have identified one ingredient among many associated with the occurrence of that mental phenomenon and that is no reduction at all. If *B* is a reductive base of a mental phenomenon, *B* cannot occur (over some non-trivial range of modal scenarios) without its associated mental phenomenon, *M*. But then *B* cannot (to the same modal degree) be mind independent. *B* is tied to *M* via the rich web of modal associations stipulated by the reductive relation. If *M* were not to exist, or if it were to have a radically different nature, *B* would not exist or have a different nature. *B* cannot be both a reductive base of *M* and be mind independent.

The puzzle should not to be confused with another puzzle about mind dependence. That puzzle arises from a worry about *causal dependence* on minds. Many entities causally depend on minds for their existence and nature: tables, chairs, cities, children. Is realism about those entities thereby threatened? (Devitt 1991; Miller 2012; Godfrey-Smith 2016). Devitt (1991) and Miller (2012) argue that it is not: a realist does not deny causal dependence on minds. Her claim is to deny a ‘further (philosophically interesting)’ sense of dependence that goes beyond ‘mundane’ causal dependence on minds (Miller, 2012).<sup>2</sup> Observe that this does not solve our puzzle. The form of mind dependence at issue for us is not causal dependence: the proposals above do not say that the reductive base *causes* mental phenomena. Removing causal dependence from the field would not help a realist. It is the constitutive mind dependence within the reductive relation that renders the anti-realist's claim trivially true.

Should we then grant the anti-realist an easy ‘win’: concede that we should be anti-realists about reductive bases in cognitive science, including neural computations? This is not an option we should contemplate. If we were to concede, anti-realism would spread to other

---

<sup>2</sup> Godfrey-Smith (2016) argues that this reply would be wrong: causal dependence on minds is relevant to realism.

entities outside cognitive science. Atoms and electrons – large collections of them – are among the (likely) reductive bases of human mental life. Collections of atoms and electrons realise (or constitute, ground, etc.) at least some human mental phenomena. The atoms and electrons occupying the space where you sit now are likely sufficient to produce (some aspects of) your mental life. If one were to replicate these atoms and electrons, one would replicate those mental phenomena. This holds true over a non-trivial range of modal scenarios. But then the puzzle can be restated for those collections of atoms and electrons. In the face of the puzzle one might, at a push, concede anti-realism about neural computation. But conceding anti-realism about atoms and electrons would be madness.

Let us see how to respond to the puzzle in a way that does not grant a win to the anti-realist.

## 5. Solutions to the puzzle

The proposals described in this section solve the puzzle by distinguishing between two types of mind dependence. Reductive theories in cognitive science entail one form of mind dependence; anti-realism about the subject matter entails another. Just because the reductive kind of mind dependence holds for cognitive science does not mean that the anti-realist kind of mind dependence holds too. The hard question is how to draw the distinction between a reductive and an anti-realist form of mind dependence. In this section, I examine three ways to do this. The first two distinguish the two kinds of mind dependence as dependence on the mind of the *subject* versus dependence on the mind of the *enquirer*. I argue that this approach is unlikely to succeed. My favoured proposal is based on attending to the structured, compositional nature of the reductive base in cognitive science. The two forms of mind dependence can be distinguished as dependence of the *component parts and relations* of the reductive base on minds (non-trivial and the point of disagreement in realist/anti-realist disputes) versus dependence of the *whole reductive base* on minds (trivial and entailed by reduction).

### 5.1 Dependence on the enquirer versus the subject

The first way to distinguish the two forms of mind dependence is to ask on *whose* mind the reductive base depends. When we described the neural computation that determines whether Abby goes to the fridge for a beer, we said that a realist should concede that Abby's neural computation depends on *her mind* but not on the mind of *anyone else*. Anti-realism about neural computations appears to be a claim about dependence on observers, not dependence on the subject being observed. Anti-realism is the claim that the world depends

on how enquirers see or conceive of it. It does not depend on how the subject being studied sees it. Anti-realism about cognitive science is in this respect no different from anti-realism about other domains. It just happens that for cognitive science the subject matter under scrutiny is the reductive base of someone's mental life. Whether this reduction obtains or not should be irrelevant to the anti-realist. Her concern is not reduction but to establish that the relevant aspect of the world depends on the mental life of others.

Drawing the distinction this way fits with the practice of cognitive science. Both the realist and the anti-realist should agree that the reductive base of some experimental subject's mental life depends on that subject's mind in the way described by the puzzle. If the experimental subject were not to have a mind, or if she were to have a radically different mind, the reductive base would be different. The reductive base is not independent of the experimental subject's own mind. But the realist and the anti-realist can disagree about whether the reductive base depends on the minds of others, specifically, on external enquirers. No justification follows for this from the reductive claim. We can state a distinction between the two kinds of mind dependence as follows. *Reductive* mind dependence is dependence on the subject's own mental life. *Anti-realist* mind dependence is dependence on the mental life of others, specifically the enquirers who study, and ascribe properties to, the reductive base.

This way of drawing the distinction handles many cases, but not all. The problem is that there is no reason to believe that two separate persons are necessary to do cognitive science. An experimental subject could, in principle, perform experiments on herself. She could provide evidence and ascribe to her own brain specific neural representations. In this case, the proposal for distinguishing two kinds of mind dependence would fail. There would not be two separate minds (subject and enquirer), so there would not be two kinds of mind dependence. Both collapse to dependence on the subject's own mind. The solution to the puzzle must lie elsewhere.

## 5.2 Dependence on second-order mental states

One might try to finesse the previous distinction by looking for a difference *within* mental life between *enquirer-like* and *subject-like* aspects. If these two aspects could be identified, we could map them onto two kinds of mind dependence. But how to draw this distinction? One plausible thought is that enquirer-like aspects are distinguishable by *being about* other aspects of mental life. A subject may have all sorts of mental states (beliefs, desires, and so on). What is special about an enquirer is that she has thoughts *about* those aspects of the subject's mental life. Enquirer-like thoughts are second-order thoughts about the mental life of a subject. The second-order thoughts might occur within a separate person

(an external enquirer) or within the same person (a subject who is her own enquirer). We therefore avoid the counterexample above of the subject who is her own enquirer. *Reductive* mind dependence would be dependence on a subject's own mental life. *Anti-realist* mind dependence would be dependence on second-order mental states, either of the subject or some other enquirer, which are about that subject's mental life.

The problem is that this proposal fails to accommodate many plausible forms of anti-realism. Consider Blackburn (1993)'s anti-realist reconstruction of Hume's view on causation. According to Blackburn, the existence and nature of causal relations depends on our cognitive apparatus – Hume is, in this sense, an anti-realist about causation. But Blackburn does not say that causation depends on our *representational* mental states, such as our beliefs or desires about causation. Causation depends on a different feature of our mental life: our dispositions to make certain inferences. Whether *A* causes *B* depends on our disposition to readily infer the occurrence of *B* from the occurrence of *A*. We have here anti-realism – anti-realism about causation – but not dependence on representational mental states.

Following this model, a form of anti-realism about cognitive science – for example, anti-realism about neural computation – need not say the relevant entities depend on *representational* mental states. Indeed, an anti-realist need not say that we have mental representations at all. She might say that neural computations depend on non-representational aspects of our mental life (for example, our dispositions to make certain inferences). Second-order mental states only make sense in the context of representational mental states. If anti-realism does not require representational states, or dependence on second-order representations, the latter notion cannot be used to distinguish anti-realism from realism. Some other form of mind dependence must be at issue.

### 5.3 Dependence of the parts versus the whole on minds

The previous two proposals try to partition the mental realm into distinct subject-like and enquirer-like parts. In certain cases, this is easy to do (for example, when subject and enquirer are two different people). But in general, it is difficult to know what distinguishes an enquirer-like aspect of the world from a subject-like aspect of the world. The current proposal adopts a different strategy. Rather than try to partition the mental realm into subject-like and enquirer-like parts, instead focus on partitions already given to us by theories in cognitive science: partitions in the reductive base.

The structured nature of the reductive base is important in cognitive science. Theories in cognitive science do not reduce a mental phenomenon to a single, undifferentiated thing, *the reductive base*. They instead reduce mental phenomena to a series of entities embedded within a structure with multiple individual parts and relations. Which parts and relations these

are varies between theory: they might be computational steps, mechanisms, networks, dynamic relations, or causal sequences of events. What does not vary is that the reductive base has some structure or another. For example, a theory that identifies an instance of decision making with a neural computation, *C*, does not reduce decision making to a single, atomic individual. Rather, the theory identifies decision making with a structured entity composed of multiple parts (including representations of environmental states, representations of utilities, and individual functional parts) and multiple relations (causal, syntactic, and other relations) that together are (or realise, constitute, ground) decision making.

Observe that, as far as our puzzle goes, only the reductive base *as a whole* is mind dependent. The entire base cannot occur without its associated mental phenomenon. But nothing follows from this regarding the mind dependence or otherwise of its *individual parts and relations*. Mind dependence of the whole reductive base does not entail mind dependence of the parts. Observe that none of the parts or relations would, by itself, be sufficient for the mental phenomenon. There is nothing incoherent in supposing that any part or relation can occur individually without any specific condition involving mental phenomena being met. Suppose that an instance of decision making is a specific neural computation. That neural computation, as we have seen, is mind dependent: it cannot occur without the associated mental phenomenon. But this does not mean that the individuals and relations that compose the computation are also mind dependent. It is possible that those individuals and relations – the representations of environment states, the functional units, the causal relations – could occur individually without any further condition being met concerning mental phenomena. It is possible that the individual parts and relations are mind independent. It is also possible that one or more of the parts and relations is mind dependent: it depends for its existence or nature on minds. Parts and relations may be mind independent even if the whole reductive base is not. Moreover, there is scope for being a different type of anti-realist by adopting anti-realism about different parts or relations within the reductive base: one might, for example, be an anti-realist about a neural computation by being an anti-realist about the relevant causal relations or by being an anti-realist about the relevant syntactic properties. In contrast, there is only one way to be a realist. A realist holds that none of the constituent parts and relations is mind dependent: each could occur individually without a further condition being met regarding mental agents or mental phenomena. Hence, we can draw our distinction. *Reductive* mind dependence is dependence of the whole reductive base on a mental phenomenon. *Anti-realist* mind dependence is dependence of one or more of the components or relations that make up the reductive base. Reductive mind dependence is entailed by the reductive claim. Anti-realist mind dependence is not.

How do we know this is the right way to draw the distinction? Recall that what is at stake is cognitive science's naturalising ambition: the attempt to show how mental life arises from non-mental ingredients. A naturalising explanation explains scientifically-relevant properties of a mental phenomenon in terms of parts and relations of the reductive base. Whether the explanation in question is functional explanation, mechanistic explanation, computational explanation, causal explanation, or another form of explanation, it invariably involves citing component parts and relations of the reductive base. The components parts and their relations explain the scientifically-relevant properties of the mental phenomenon. As we defined it above, an explanation is naturalising if its *explanans* does not refer to, or otherwise presuppose, mental phenomena. The relevant explanations in cognitive science appeal to the individual parts and relations of the reductive base and not to the structure as a whole. If we are realists about those parts and relations, we can appeal to them without presupposing further conditions being met about mental phenomena. If the naturalising project succeeds, we must be realists about the relevant parts and relations referred to in our explanations. Conversely, if one or more of the parts or relations that make up the reductive base are mind dependent, then an explanation that cites them will not naturalise the relevant mental phenomenon. The distinction we draw coincides with what matters to the naturalising project.

Consider an analogy. You see a miniature castle in a shop window. You want to explain some of the castle's properties: why it can bear so much weight or why it is resistant to attack by crumpled-up paper balls. You want your explanation to 'naturalise' the castle: to explain the castle's properties in non-castle-dependent terms. You do not want your explanation to make reference to, or otherwise presuppose, castles. Closer inspection reveals that the castle is built from Lego bricks. You advance a reduction: the castle is (or is realised by, or is constituted by) this configuration of Lego bricks. Armed with this reduction, you can explain the effects first noted. The individual Lego bricks and their specific configuration explain the ability of the castle to bear weight. The individual Lego bricks and their configuration explain the resistance of the castle to attack by paper balls. Someone might object that, according to your hypothesis, the castle is (or is realised by, constituted by) this configuration of Lego bricks. Hence, you have not explained the castle in non-castle-dependent terms. You reply, rightly, that this kind of castle dependence would be orthogonal to your naturalistic ambitions. The giant configuration of Lego bricks is a castle, but the individual bricks and their relations are not. Your *explanans* cites those individual bricks and their relations not the configuration as an atomic whole. You have explained weight-bearing and resistance-to-attack in terms of these parts and relations, neither of which are castles or are castle-dependent. That is all that is required to naturalise the castle. Now suppose that we

discover, to our surprise, that individual Lego bricks essentially depend on castles. Perhaps those Lego bricks must contain tiny castles. The configuration of Lego bricks is now castle dependent in a new and more troublesome way. The original naturalising ambition – explaining the castle’s weight-bearing and resistance-to-attack without making reference to, or otherwise presupposing, castles – would have failed.

## 6. Conclusion

I have argued that what matters to the realist/anti-realist dispute in cognitive science is not whose mind the reductive base depends on (subject versus enquirer). Rather, what matters is the mind dependence of the individual parts and relations versus that of the reductive base as a whole. The status of the individual nuts and bolts that realise cognition matters. Whether that collection of nuts and bolts taken as a whole is mind dependent is irrelevant. Perhaps surprisingly, the structured nature of reduction in cognitive science, and cognitive science’s parallel emphasis on explanation of mental phenomena in terms of parts of that reductive base (whether that be functional, mechanistic, computational, causal, or another form of explanation) turns out to be important to articulating the realist/anti-realist dispute in this area. The relevant form of realism in this context cannot be disentangled from reduction to compositional structures. If someone claims to be an anti-realist about cognitive science, one should ask: about which entities are you anti-realist and do those entities play an essential role in the reductive base of cognition?

## References

- Blackburn, S. 1993. “Hume and Thick Connexions.” In *Essays in Quasi-Realism*, 94–107. Oxford: Oxford University Press.
- Block, N. 2007. “Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience.” *Behavioral and Brain Sciences* 30: 481–548.
- Burge, T. 1986. “Individualism and psychology”. *Philosophical Review* 95: 3–45.
- Churchland, P. M. 1981. “Eliminative Materialism and the Propositional Attitudes.” *The Journal of Philosophy* 78: 67–90.
- Dennett, D. C. 1991a. *Consciousness Explained*. Boston, MA: Little, Brown & Company.
- . 1991b. “Real Patterns.” *The Journal of Philosophy* 88: 27–51.
- Devitt, M. 1991. *Realism and Truth*. 2nd ed. Princeton, NJ: Princeton University Press.
- Fodor, J. A. 1975. *The Language of Thought*. Sussex: The Harvester Press.
- Fodor, J. A. 1980. “Methodological solipsism considered as a research strategy in cognitive psychology.” *Behavioral and Brain Sciences* 3: 63–109.



- Godfrey-Smith, P. 2016. "Dewey and the Question of Realism." *Noûs* 50: 73–89.
- Gold, I. J., and M. N. Shadlen. 2001. "Neural Computations That Underlie Decisions About Sensory Stimuli." *Trends in Cognitive Sciences* 5: 10–16.
- . 2007. "The Neural Basis of Decision Making." *Annual Review of Neuroscience* 30: 535–74.
- Irvine, E. 2012. *Consciousness as a Scientific Concept*. Dordrecht: Springer.
- Miller, A. 2012. "Realism." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Spring 2012. <http://plato.stanford.edu/archives/spr2012/entries/realism/>.
- Putnam, H. 1988. *Representation and Reality*. Cambridge, MA: MIT Press.
- Rangel, A., C. Camerer, and P. R. Montague. 2008. "A Framework for Studying the Neurbiology of Value-Based Decision Making." *Nature Reviews Neuroscience* 9 (545–556).
- Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson.
- Schultz, W., P. Dayan, and P. Read Monague. 1997. "A Neural Substrate of Prediction and Reward." *Science* 275: 1593–9.
- Searle, J. R. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Shoemaker, S. 2007. *Physical Realization*. Oxford: Clarendon Press.
- Sperling, G. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs: General and Applied* 74: 1–29.
- Wilson, R. A. 2001. "Two Views of Realization." *Philosophical Studies* 104: 1–30.