

Realism about cognitive science

Mark Sprevak
University of Edinburgh

15 November 2017

This chapter starts from a puzzle. Realism about X is often glossed as the idea that X s are *mind independent*: X s exist, and have their nature, independently of our beliefs, interests, attitudes, or other mental states. X s are, in a sense ‘out there’, getting on with it independently of our mental life. If this is right, how should we understand realism about cognitive science? Mental processes and states are not mind independent: they don’t take place independently of our beliefs, interests, attitudes, or other mental states. Hence, it seems that we cannot be realists about them. Nevertheless, in line with other areas of philosophy of science, there seems scope for asking the realist question about the posits of cognitive science even if those posits make up our mental life. But unless we state realism differently, there is no way to sensibly ask the realist question about cognitive science. In this paper, I explore the right way to state realism about cognitive science. I introduce three different types of mind dependence and evaluate their merit to stating realism about cognitive science.

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, independently 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. But how then can one be a realist about them? This is our puzzle. My solution will be to distinguish between two types of mind dependence in 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 varieties of realism that one might adopt about cognitive science. In Section 3, I argue that realism that asserts mind independence has a special role to play in cognitive science. In Section 4, I present the puzzle about this variety of realism. In Section 5, I examine three solutions to the puzzle. Each draws a distinction between a trivial and a non-trivial form of mind dependence in a different way. My favoured proposal derives from the observation that theories in cognitive science aim to explain mental phenomena in terms of structured complexes – for example, in terms of 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’ versions of realism should also be distinguished from ‘global’ versions. A global version of realism would assert realism about all or most subject matters of the mature sciences. I will not consider global versions of realism here. My concern is with realist claims made about entities specifically in cognitive science.

Within this domain, a realist may make a range of claims. Realist/anti-realist disputes take on a different character depending on which claim is at issue. In this section, I highlight six possible varieties 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. A realist may assert or deny these claims in various combinations.

First, *existence*. On this view, realism about *Xs* commits one to the existence of *Xs*. 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? This second variety of realism holds that Xs are discrete individuals. Fodor (1975) is a realist in this sense about beliefs. Beliefs are discrete individuals that occur and re-occur inside someone's head. The relevant kind of anti-realism would take a deflationary view of the relevant entity. Dennett (1991b) argues that beliefs are not discrete individuals but rather amorphous and hard-to-count 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 a realist in this sense too 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, that thing is a tokening of a sentence in the language of thought 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 that Abby has. Instead, we intend to convey to our listeners 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 of subjective phenomenal aspects of their experience (reporting, for example, that they experience red). We should, according to Block, understand these reports as aiming to tell the truth about those experiences. 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 of phenomenal aspects of experience but as a roundabout way for the subject to express that her cognitive system has detected a highly disjunctive physical property (such as redness). A realist holds that the discourse aims at telling the truth. An anti-realist denies the truth-seeking character of discourse but may maintain that the talk has other virtues (e.g. pragmatic virtues).

Fifth, *evidence*. If one is a realist about Xs, then one holds that we have justification for the truth (or approximate truth) of the relevant part of the discourse. Block (2007) is a realist in this sense too about phenomenal consciousness. Subjects' reports of conscious experience not only aim to tell the truth about instantiations of phenomenal character, we also (normally) have justification that they *are* true. Significantly, the justification holds even under unusual presentational conditions such as when stimuli are flashed 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 some degree of epistemic caution about the relevant claims. Irvine (2012) claims that we lack justification for believing the reports of subjects about phenomenal aspects of their experiences in the context of Sperling's experiments.

Finally, *mind independence*. Like the second claim, this concerns the nature of the entities. However, the question here is not about their nature as discrete individuals but about their degree of mind dependence: does that entity depend, for its existence or nature, on minds? All our knowledge of the world is mediated to some extent by our minds. We cannot see the world untouched 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 nature. But seeing the world 'as it is', without some contribution from the human mind, is impossible. This invites a question: Which parts of our knowledge represent to entities and properties that are really out there and which are (partial) constructions of our minds? Some entities appear to exist and have the properties that we attribute to them independently of the way we think of them. Perhaps some fundamental particles in physics, e.g. electrons, are like this. If our minds were not to exist, or if they were to have a radically different nature, electrons would continue to exist and have unchanged properties. Other entities appear to be partial constructions of our minds. Beauty marks may be an example of these. Whether a specific skin colouration is a beauty mark depends on how that colouration strikes, or would strike, a mind like ours and fit with our visual preferences – whether that patch *looks beautiful to us*. If human minds were not to exist, or if they were to have a different makeup, the distribution of beauty marks in the world would be different. One might ask the realism/anti-realism question about the entities of cognitive science. For example, among those entities are neural computations. Neural computations are invoked by cognitive science 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 subject concerned performs specific neural computations. Cognitive science invokes neural computations to explain mental life. Should one be a realist or an anti-realist about these neural computations?

Fodor (1980) is an example of someone who is a realist about these neural computations. 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 a 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 is determined by facts about Abby and her brain. Burge (1986) is another realist about neural

computation but he holds that the neural computation depends on a broader base of mind-independent facts: it depends not just on Abby's brain but also her causal relationship to her environment. Despite their disagreement, both Fodor and Burge agree that neural computations are really out there, they are not a grouping that is dependent on how we human agents view Abby – a grouping that is somehow significant to us but not reflective of any objective distinction in the world. The aim of computational cognitive science is to discover and describe these objective distinctions in the world carved out by neural computations. One may get this description right or wrong, but one does 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 them, neither Abby's brain nor her brain plus her relation to her environment determine whether her brain performs a specific computation. Absent consideration of *how we view* Abby, there is no fact about whether Abby's brain performs one computation rather than another, or whether it performs any computation at all. Neural computations are observer relative. If human minds were not to exist, or if they were to have a different makeup, the distribution of neural computations would be different. Neural computations are more like beauty marks than electrons: they are a construction that reflects the specific way in which humans are disposed to conceive of the world, not objective features waiting 'out there' to be discovered.

For electrons and beauty marks, the question about mind dependence can be posed in a relatively straightforward manner. The worry is that the same cannot be said for neural computations. Neural computations are hypothesised to be connected to mental life. They realise or otherwise constitute aspects of our mental life. This makes realism about neural computations hard to understand as a coherent possibility. If neural computations realise mental life, how can neural computations be mind independent? Fodor and Burge cannot believe that Abby's neural computations are entirely mind independent. The entity in question – the neural computation that underlies Abby's decision making about the beer – depends on at least one mind: Abby's own. If Abby's mind were not to exist or to have a different nature, that neural computation would differ. Similarly, Putnam and Searle cannot believe that Abby's neural computations are *in some way or other* mind dependent. That would be trivially true. No one thinks that her neural computations can exist, or have their nature, independently of how things go in Abby's mental life. So both the realist and the anti-realist must agree that Abby's neural computations are mind dependent *in some way or other*. The realist/anti-realist dispute cannot therefore be about mind dependence *simpliciter*. Something else must be going on. Identifying what this is – what is at stake in this realist/anti-realist dispute in cognitive science – is our puzzle.

Earlier in this section we saw that a realist about X need not endorse a mind independence claim about X . We saw five alternative ways to be a realist about cognitive science. This suggests a quick way out of our puzzle. If stating mind independence is a problem for cognitive science, why not simply abandon this form of realism and pursue some other form of realism? There are at least five other options to choose from. In the next section, I argue that while there is nothing wrong with these alternative other forms of realism about cognitive science, this strategy would have a significant cost. Cognitive science needs the mind-independence form of realism to fulfil one of its wider ambitions: the ambition to naturalise the mind.

3 Why care about mind independence?

The world contains at least two kinds of phenomenon: mental phenomena – involving entities like beliefs, sensations, ideas, concepts, thought processes, judgments, and so on – and physical phenomena – involving entities like 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 just because 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 reductively pair specific mental phenomena with non-mental phenomena. The non-mental phenomena often have specific properties: the states *perform computations*, *represent*, *process information*, *carry error signals*, and so on. Certain instances of decision making, for example, are paired with certain 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 reliably correlate across a wide range of circumstances including conditions not experimentally tested. Precisely how far this modal dimension extends – across every possible world, across worlds with the same physical laws as ours, across worlds with the same natural laws as ours – is open to question, but we can be sure that the association has a non-trivial modal dimension. The second observation is that the non-mental member of the relationship could substitute for its mental counterpart

without change in scientifically relevant effects. For example, the scientifically relevant effects 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 effects. The drift-diffusion model, for example, aims to provide not just a neural correlate of decision making but also to show that this correlate would produce the characteristic effects 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 modal circumstances and it also generates all the scientifically relevant effects associated with that mental phenomenon, then we are in a position to advance a reductive claim. 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 entities bear some reductive relation – perhaps *identity*, *realisation*, *constitution*, *grounding*, or another relation – 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* a neural computation.

The theories in question identify some kind of reductive base for a mental phenomenon. The details of the reductive relation may differ (*identity* vs. *realisation* vs. *constitution* vs. *grounding*). But the general idea of finding some non-mental base that is sufficient for the scientifically relevant 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, and somehow produces, the scientifically relevant properties of the mental phenomenon.

Successful reductions of this sort appear to provide a road to naturalising the mind. By ‘naturalising’ I mean explaining scientifically relevant effects of mental phenomena in non-mental terms: in terms of a subject matter that does not already presuppose mental life. A naturalising explanation is one that takes as its *explanandum* some scientifically relevant effect 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 of the brain). Naturalising the mind therefore requires realism about the subject matter of the *explanans*. One needs to be a realist – in the sense of asserting mind independence – about the entities cited by the *explanans*. To see why consider the alternative. Suppose that anti-realism about neural computation is correct. Explain-

ing decision making by appeal to neural computation would in this case not serve to naturalise that mental phenomenon. Explanation in terms of neural computations would not explain the phenomenon in non-mental terms. It would explain the phenomenon in terms of entities that depend on minds for their existence and nature. 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 naturalising. One might of course still explain decision making in terms of neural computation. But one should not think that this provides a way to naturalise the mind: one has not shown how decision making arises from non-mental ingredients. Rather, one has offered a non-reductive explanation: an explanation of a mental phenomenon in terms, *inter alia*, of other mental phenomena. Nothing wrong with this – *per se*. But it does not serve an ambition to naturalise the mind. In order for the naturalising strategy described above to work we need to be realists – specifically, realists who assert mind independence – about the subject matter of our *explanans*.

Realism about the subject matter of cognitive science is not a mere idle intellectual posture. Realism of the mind-independence variety is needed for explanations within cognitive science to serve the project of naturalising the mind. It is perfectly possible to pursue cognitive science without any naturalistic ambition. But giving up that ambition is not to be taken lightly. Consider what we would miss out on: understanding how the mind arises from non-physical ingredients. Rather than abandon this variety of realism, let us instead examine what the problem with it is and 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 appears to open the door to naturalising the mind. However, this can only work if one can be a realist – in the sense of asserting mind independence – about the non-mental side of the relation. The problem is that the preceding two claims – (i) mental phenomena reduce to non-mental phenomena; (ii) the non-mental side of the relation is mind independent – are incompatible. This is our problem. Let us examine it in more detail.

Consider what happens if the reductive relation in question is *identity*. Assume that some instance of human decision making *is* a specific neural computation. In order to use this to offer a naturalistic explanation of 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 the neural computation be

mind independent? If decision making *is* a neural computation, then that neural computation must be mind dependent. Being identical to a mental phenomenon surely entails mind dependence. If human minds were not to exist, or if they were to have a different makeup, the existence and nature of the neural computation would be different. What stronger reason could there be for thinking that *X* is mind dependent than *X* being *identical* to a mental phenomenon? But if the neural computation is mind dependent, then anti-realism is true and realism is false. The reduction of decision making to a specific neural computation seems to preclude realism about that neural computation.

What if the reductive relation in question were *realisation*? Identity is a symmetric relation: if *X* is identical to *Y*, then *Y* is identical to *X*. Perhaps it is the symmetry of this reductive relation that is the source of the problem. Realisation is asymmetric: if *X* realises *Y*, *Y* does not realise *X*. Would an asymmetric relation allow us to avoid mind dependence of one side of the relation infecting the other side? Unfortunately, no. The reason is that in spite of realisation being an asymmetric relation the reductive base still cannot occur *independently* of its mental phenomenon, which is what the realist requires. Suppose that an instance of decision making is realised by a neural computation. If it is realised by that neural computation, then that neural computation is sufficient for that instance of decision making to occur.¹ The occurrence of that neural computation is sufficient for the occurrence of that instance of decision making; otherwise, it would be unclear why what we found was a reductive base at all. So the following conditional holds: *If* this neural computation were to occur, *then* the relevant decision-making process would occur too. Moreover, this conditional holds over a non-trivial range of modal circumstances (the precise extent is determined by the realisation relation in question). The neural computation is tied to the mental process in a modally rich way such that the computation cannot occur without the relevant decision making also occurring. But then the neural computation is not mind independent. The neural computation cannot occur independently of minds. It cannot occur without the associated mental process also occurring. The reductive base is not mind independent. Let us put the same point schematically. Suppose that a reductive base, *B*, realises some mental process, *M*. *B* is tied in the modally rich way entailed 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 without *M* and hence *B* cannot occur without specific mental phenomena occurring. If human minds were not to exist, or if they were to have a different makeup (e.g. without *M*), then the facts about *M* would be different so the facts about *B* would be different.

¹I 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).

The realist's claim that *B* is mind independent is flat out incompatible with the claim that *B* realises *M*.

Other reductive relations – grounding or constitution – suffer from the same problem. The reason is that for any reductive relation, the reductive base should, in some modally rich sense, be *sufficient* for the mental phenomenon. It should 'bring about' that mental phenomenon. The specific content of 'bringing about' will be cashed out in different ways by different reductive relations. Irrespective of differences between reductive relations the reductive base must be sufficient for the mental phenomenon – otherwise, why think we have identified a reductive base at all? If the 'base' is not sufficient for the mental phenomenon, then we have identified only one ingredient among (possibly many) others 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 circumstances) without the associated mental phenomenon, *M*. But then *B* cannot (to the same modal extent) be mind independent. *B* is tied to *M* via the web of associations stipulated by the reductive relation. If *M* were not to exist, or if it were to have a different nature, *B* would not exist or it would have a different nature. *B* cannot be both a reductive base of *M* and be mind independent.

The puzzle should not be confused with a similar puzzle about mind dependence. That puzzle arises from a worry about trivial *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 undermined (Devitt, 1991; Miller, 2012; Godfrey-Smith, 2016)? Devitt (1991) and Miller (2012) argue that it is not because a realist does not deny causal dependence on minds. Anti-realism, rather than asserting causal dependence, is defined by a 'further (philosophically interesting)' sense of dependence that goes beyond 'mundane' causal dependence on minds (Miller, 2012).² This does not help us with our puzzle. The form of mind dependence at issue for us is not causal dependence. The proposals above do not say that reductive bases *cause* mental phenomena. Removing causal dependence from the field would not help us here. It is the further, non-mundane, 'constitutive' mind dependence built into the reductive relation that renders the anti-realist's claim about cognitive science trivially true.

In response to the puzzle, should we then grant the anti-realist an easy 'win': concede that we should be anti-realists about the reductive base of mental phenomena in cognitive science, including neural computations? This is not an option we should contemplate. If we were to concede to the anti-realist here, anti-realism would spread to other entities outside cognitive science. Atoms and electrons – large

²Godfrey-Smith (2016) argues that this reply is wrong: causal dependence on minds is relevant to realism.

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 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 conditional holds true over a non-trivial range of modal scenarios. But then the argument of this section can be applied to these collections of atoms and electrons. At a push, one might concede anti-realism about neural computation. But conceding anti-realism about atoms and electrons on the basis of the argument above seems 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

Each of the proposals described in this section solve the puzzle by distinguishing between two types of mind dependence. Reductive theories in cognitive science involve one form of mind dependence whereas anti-realism about the subject matter involves another. The hard question is how to draw the distinction between a (trivial) reductive form of mind dependence and a (non-trivial) 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 based on dependence on the mind of the *subject* versus dependence on the mind of an *enquirer*. I argue that this approach is unlikely to succeed. My favoured proposal is based on attending to the structured 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 a neural computation that determines whether Abby goes to the fridge for a beer, we said that Abby's neural computation trivially depends on *her* mind but we did not comment on whether it depends on the mind of *anyone else*. One might propose that anti-realism about neural computations is a claim about dependence on observers, not a claim about

dependence on the subject being observed. Anti-realism in general 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. Cognitive science appears to be special only in that the subject being observed has a mind. Whether the subject has a mind or not should be irrelevant to the anti-realist. Her concern is not to establish dependence of the subject on her own mind but to establish dependence of the subject on the mental life of others.

Drawing the distinction this way also fits with the practice of cognitive science. Both the realist and the anti-realist can 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. But the realist and the anti-realist can disagree about whether the reductive base depends on the minds of external enquirers. No justification for this flows from the reductive claim. We can state our distinction between 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 proposal by looking for a difference *within* a subject's mental life between her *enquirer-like* and *subject-like* aspects. If these two aspects could be identified, we could map them onto our two kinds of mind dependence. But how to draw this distinction? One thought is that enquirer-like aspects of mental life are distinguished 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 her enquirer-like thoughts is that they are *about* aspects of her 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. On this view, *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's characterisation of anti-realism fails to fit many plausible forms of anti-realism. Consider Blackburn (1993)'s anti-realist reading 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 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 anyone's 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). The distinction between first and second-order mental states only makes sense in the context of representational mental states. If anti-realism does not require representational states, the first-order/second-order distinction 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 mind into subject-like and enquirer-like parts. In certain cases, this may be feasible (for example, when subject and enquirer are in two different people). But in general, it is difficult to know what distinguishes an enquirer-like aspect of the mental world from a subject-like aspect of the mental world. The proposal in this section adopts a different strategy. Rather than try to partition the mental realm into subject-like and enquirer-like parts, let us instead attend to partitions already given to us by theories in cognitive science: partitions in the reductive base.

The structured nature of the reductive base is important to cognitive science. Theories in cognitive science do not reduce a mental phenomenon to a single, undifferen-

tiated entity. They reduce mental phenomena to a structured entity that consists of multiple individual parts and relations. Which parts and relations these are varies between theories: they might be computational steps, mechanisms, networks, dynamic relations, or causal sequences of events. 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, *C*. Rather, the theory identifies decision making with a structured entity, *C*, composed of multiple parts (perhaps 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 the puzzle described in the previous section only entails that the reductive base *as a whole* is mind dependent. The reductive base cannot occur without its associated mental phenomenon. But nothing follows from this regarding the mind dependence of the *individual parts and relations*. Mind dependence of the whole reductive base does not require mind dependence of the parts. The same argument as above cannot be run for the parts as there is no reason to suppose that any of the individual parts or relations would, by itself, be sufficient for a mental phenomenon. There is nothing contradictory in supposing that a part or relation of the reductive base can occur individually without any specific condition involving mental agents being met. For example, suppose that an instance of decision making is a specific neural computation. That entire neural computation is mind dependent: it cannot occur without the associated mental phenomenon. But this does not mean that the individual parts and relations that compose the computation are also mind dependent. It is possible that the individual parts and relations – the representations of environment states, the smaller functional units, the causal relations – could occur individually without any condition being met concerning mental agents. It is also possible that one or more of the parts and relations is mind dependent. Parts and relations may be mind dependent or fail to be mind dependent even if the whole reductive base is not mind dependent. There is scope for different anti-realist views by adopting the mind-dependence claim about different parts or relations: one might, for example, be an anti-realist about causal relations or about syntactic properties. By contrast, there is only one way to be a realist: hold that none of the constituent parts and relations is mind dependent. Each part or relation could occur individually without a further condition being met regarding mental agents. 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 constituent individual parts 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 our distinction? Recall that what is at

stake is the ambition to naturalise the mind: the attempt to show how mental life arises from non-mental ingredients. A naturalising explanation explains properties of a mental phenomenon in terms of the individual parts and relations of the reductive base. Whether the form of explanation in question is functional explanation, mechanistic explanation, computational explanation, causal explanation, or some other form of explanation, it consists in citing the individual parts and how they are arranged by relations in the reductive base. The individual parts and their relations explain the scientifically relevant properties of the mental phenomenon. As we defined it above, an explanation is naturalising only 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. If we are realists about those parts and relations, we can appeal to them in our explanations without presupposing further conditions being met about mental phenomena. Conversely, if the naturalising project is to succeed, we must be able to be realists about the relevant parts and relations referred to by our explanations. In contrast, 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 fail to naturalise the mental phenomenon. That the proposed distinction aligns with the fate of our naturalising ambitions indicates that we are on the right track here.

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 aim for your explanation to be 'naturalistic': to explain the castle's properties in non-castle-involving terms. You do not want that explanation to make reference to or otherwise presuppose castles. Closer inspection reveals that the castle is built from Lego bricks. You make a reductive claim: the castle is (or is realised by, or is constituted by) this specific configuration of Lego bricks. Armed with this reductive claim, 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 at this point that, according to your hypothesis, the castle is this configuration of Lego bricks. Hence, you have not explained the castle in non-castle-involving terms. You reply, rightly, that this kind of castle involvement does not matter to your naturalistic ambitions. The specific 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 the powers of these parts and relations, neither of which are castles or are castle-dependent. That is all that is required to naturalistically explain the castle's properties. Now suppose that one were to discover, to great surprise, that the individual Lego bricks

do essentially depend on castles. Perhaps some Lego bricks contain tiny castles. The structured 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 fail.

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, it is the mind dependence of the individual parts and relations versus the (trivial) mind dependence of the reductive base as a whole. The status of the individual nuts and bolts that realise cognition matters. Whether a specific configuration of nuts and bolts taken as a whole is mind dependent is irrelevant to realism and to the naturalising project. Perhaps surprisingly, the structured nature of the reductive base in cognitive science and cognitive science’s parallel emphasis on structured explanation (whether that be functional, mechanistic, computational, causal, or another form of explanation via appeal to a structure) turns out to be essential to articulating the realist/anti-realist dispute in this area. The relevant form of anti-realism targets one or more entities in that structure. If someone claims to be an anti-realist about cognitive science, the first question one should ask is: About which entities in the reductive base are you anti-realist? The next question should aim to discover whether those entities really do play an essential role in the reductive base of cognition.

Bibliography

- Blackburn, S. (1993). “Hume and thick connexions”. In: *Essays in Quasi-Realism*. Oxford: Oxford University Press, pp. 94–107.
- Block, N. (2007). “Consciousness, accessibility, and the mesh between psychology and neuroscience”. In: *Behavioral and Brain Sciences* 30, pp. 481–548.
- Burge, T. (1986). “Individualism and psychology”. In: *Philosophical Review* 95, pp. 3–45.
- Churchland, P. M. (1981). “Eliminative materialism and the propositional attitudes”. In: *The Journal of Philosophy* 78, pp. 67–90.

- Dennett, D. C. (1991a). *Consciousness Explained*. Boston, MA: Little, Brown & Company.
- (1991b). “Real patterns”. In: *The Journal of Philosophy* 88, pp. 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.
- (1980). “Methodological solipsism considered as a research strategy in cognitive psychology”. In: *Behavioral and Brain Sciences* 3. (Reprinted in D. M. Rosenthal, editor, *The Nature of Mind*), pp. 63–109.
- Godfrey-Smith, P. (2016). “Dewey and the question of realism”. In: *Noûs* 50, pp. 73–89.
- Gold, I. J. and M. N. Shadlen (2001). “Neural computations that underlie decisions about sensory stimuli”. In: *Trends in Cognitive Sciences* 5, pp. 10–16.
- (2007). “The neural basis of decision making”. In: *Annual Review of Neuroscience* 30, pp. 535–574.
- Irvine, E. (2012). *Consciousness as a Scientific Concept*. Dordrecht: Springer.
- Miller, A. (2012). “Realism”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Spring 2012. URL: <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”. In: *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”. In: *Science* 275, pp. 1593–1599.
- 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”. In: *Psychological Monographs: General and Applied* 74, pp. 1–29.
- Wilson, R. A. (2001). “Two views of realization”. In: *Philosophical Studies* 104, pp. 1–30.