

Functionalist Interrelations Amongst Human Psychological States *Inter Se*, ditto for Martians

Nicholas Shea

Forthcoming 2019 in Smortchkova, J., Dolega, K., Schlicht, T. (eds.)
What Are Mental Representations? (Oxford: OUP)

Abstract

One job for theories of mental representation is to distinguish between different kinds of mental representation: beliefs, desires, intentions, perceptual states, etc. What makes a mental state a belief that *p* rather than a desire that *p* or a visual representation that *p*? Functionalism is a leading approach for doing so: for individuating mental states.

Functionalism is designed to allow that psychological states can be multiply realized. Mark Sprevak has argued that, for a functionalist account of psychological states to apply to creatures that are organised in a very different way to humans (call them Martians), the way a psychological state is functionally individuated has to be relatively coarse-grained (Sprevak 2009). Psychological research might show that human beliefs are directly available to consciousness, that they are formed as the result of deliberate judgement, and so on, but theorists would be precluded from including these roles in their account of belief, if Sprevak is right.

The argument for coarse-grained individuation fails if we distinguish functionalism about what it takes to be a psychological state in general from functionalism about a particular state type such as belief. Functionalism individuates a psychological state like believing that *p* partly by reference to its relations to other psychological states: desiring that *p*, perceiving that *p*, intending that *p*, etc. Functionalist motivations do indeed suggest that Martians with a functional organisation and physical substrate quite unlike humans could have psychological states, but not that they should have states with the interrelated collection of functional roles to count as beliefs, desires, intentions, etc. Thus, theorists are not precluded from including functional relations to consciousness or deliberate judgement in their account of (human) belief, consistent with allowing that Martians would have their own collection of functionally interrelated psychological states.

Sprevak's coarse-grained functionalism implies an implausibly liberal form of extended cognition. The point about functional interrelations allows us to avoid that conclusion without jettisoning functionalism (as

Sprevak suggests we should): records in a human notebook may not enter into the right interrelations with other human psychological states to count as beliefs; nor do they enter into any interrelations with Martian psychological states. Functionalism can therefore allow that Martians have psychological states while holding that few if any of the beliefs we humans have are, as a matter of fact, extended.

Keywords

functionalism, commonsense functionalism, psychofunctionalism, belief, extended mind, extended cognition

Word count: 4,500

(1) Introduction

How fine-grained or coarse-grained should functionalist specifications of mental states be? Fineness of grain is a matter of how many and various the clauses are that appear in a functionalist specification of a particular mental state type. Both commonsense functionalism and psychofunctionalism admit of more and less fine-grained variants. A very fine-grained version of functionalism could have the consequence that no actual organisms other than humans have beliefs. That would still be compatible with one functionalist intuition, namely that psychological categories are neutral about the substrate in which they are realised. However fine-grained, a functional category could in principle be realised in a substrate other than carbon.

There is, however, a second prominent motivation for functionalism – a motivation that does seem to be undermined by individuating mental states in a very fine-grained way. The intuition is that we ought not to be too anthropocentric in characterising what it takes to have psychological states. That motivates functionalism, because functional specifications are less parochial and more apt to be realised by a range of different organisms. The octopus, which engages in complex and seemingly intelligent behaviour, has a very different way of organising information and generating behaviour (Godfrey-Smith 2016). The hypothetical intelligent Martian could be just as different from humans in its way of organising information and generating behaviour, and while also being based on a different substrate. Nevertheless, we should not rule out that octopuses and Martians have mental states. This motivation is thought

to suggest that functionalist specifications should be relatively coarse-grained, so that they can apply to creatures that are very different from us.

The second motivation for functionalism has played an important role in underpinning the extended mind hypothesis (Chalmers 2008, Clark 2008, pp. 88-9). It also has wider significance, because it raises the question of how we ought to individuate psychological states in general and beliefs in particular. Even if our primary interest is not in the possibility of cognitive extension, the broader question of how to individuate beliefs turns partly on whether it is appropriate to individuate them in a coarse-grained or a fine-grained way.

Within the extended mind literature, Clark has argued that the claim that mental states extend into the world beyond the skin is based on functionalist specifications of cognitive states like belief being relatively coarse-grained (2007, p. 167). He also argues that this way of individuating mental states is supported by commonsense psychology (2008, p. 88). Sprevak has turned these arguments into a purported reductio of functionalism. He argues that functionalism implies an unacceptably liberal form of cognitive extension (Sprevak 2009). Any functionalist treatment of mental states that is coarse-grained enough to apply to Martian psychology, as functionalists intend, will also count as cognitive the information contained in many of the artefacts with which humans interact, like the hard disk of your computer.

This paper questions that argument. Section (2) argues that beliefs are individuated by reference to other human psychological states. Section (3) sets out the kind of Ramsey sentences that define human psychological states and those that define Martian psychological states. A resource that would count as psychological because of its relation to other Martian psychological states, for example an external notebook, does not thereby qualify as falling under any psychological category when a human interacts with it. The anti-parochialist motivation for functionalism is therefore consistent with psychological states being individuated in a relatively fine-grained way. I conclude that the argument that functionalism entails radical extension fails.

(2) Functionalism Connects Belief with other Human Psychological States

Some have argued that there are conditions on a state's being a belief which limit the ambit of cognitive extension.¹ For example, maybe beliefs have to be readily and fluidly integrated with one another so as to allow continuous checking for consistency, in a way that semantic memories are but entries in notebooks are not. Such additional constraints would foreclose the problematic consequence that every informational resource with which a human interacts is part of her extended mind (Rupert 2004, 2009). Sprevak argues that that move is unavailable to the functionalist who is motivated by the intuition that Martians too have psychological states, since such fine-grained conditions would not be satisfied by Martian mental states. Sprevak rejects radical extension and embraces the *modus tollens*: functionalism about cognitive states is untenable.

However, it is important here that we distinguish what it is to be a psychological state in general from what it is to be a psychological state of a particular kind, like a belief. Humans have beliefs, desires, hopes and intentions; visual perceptions, auditory perceptions and sensations like pain; anger, happiness, fear; and so on. One central concern of psychology is to characterise each of these psychological states; that is, to say how they are to be individuated. The subject matter, then, is the psychological states found in humans. This is not to presuppose that only humans can have them. It is an open empirical question whether other animals share any psychological states with humans – a question which turns in part on how such states are to be individuated. It could turn out that many other animals have beliefs, and that other primates have concepts, say. The target of this enquiry is the collection of psychological states that we humans happen to have, not those (if any) that are proprietary to humans.

Another important, less empirical, question asks what it takes for there to be mental or psychological states in general. There the target is not the class of psychological states that humans happen to have, but the class of possible psychologies, which may be much wider. Possible psychology includes other ways that an organism could be set up to have some psychological organisation or other. In addressing that question we should not presuppose that the inventory of mental states found in humans is the only possible psychological organisation.

¹ Clark and Chalmers (1998); from the opposite perspective: Adams and Aizawa (2001, 2008).

Maybe there are quite different ways of processing information so as to respond flexibly and intelligently to the environment.

What does it take to have some form or other of psychology, cognition or mental life? Considerations that seem relevant include that the organism engages in consistent and complex forms of behaviour that are responsive to features of its environment; that it pursues and achieves goals that are relevant to its interests; and that it processes information in rational or intelligent ways. The question is related to debates about the 'mark' of the mental, although there the focus is the form of psychological life exemplified by humans. To disambiguate we can use 'psychological' as the more general term and reserve 'mental' for the categories of human psychology. Consciousness is a candidate but contentious mark of the mental but it is unlikely all states that count as psychological are conscious. Intentionality is also contentious in this respect, both because it is unclear whether all human psychological properties exhibit intentionality (e.g. moods, emotions on some views), and because many prima facie non-psychological artefacts seem to have intentionality. These are difficult questions, and it is in any event not the aim of this paper to give an account of what it takes to have psychological states. The important claim is that the class of possible psychological states may be much wider than the collection of psychological states that humans happen to have (even granting that some human psychological states may turn out to be instantiated more widely, for example in other animals on Earth, with which we share common ancestors).

A particular kind of mental state, like a belief, is individuated by its functional relations to stimuli, behaviour and other psychological states. A plausible constraint is that the 'other psychological states' figuring in a functionalist specification should be drawn from the same kind of psychology as the state in question; for humans, from the inventory of other human psychological properties. So human beliefs will be type-identified partly in terms of their relations to human desires and human intentions. Similarly, Martian mental states will be individuated by reference to Martian mental state types. Although it is widely recognised that functionalist specifications should include relations to other mental states – that's what sets functionalism apart from behaviourism – a potential restriction to states from the same psychological system (either human or Martian, not both) is less widely noted.

A functionalist can hold that a Martian has psychological states without being committed to the Martian's having beliefs (or any of the other types of mental states

instantiated by humans). Human mental states are individuated by a set of functional relations between perception, memory, belief, desire, intention, language processing, and so on. Martian mental states are individuated by a set of functional relations amongst the (different) collection of categories that characterise Martian psychology. In both cases the constitutive interrelations may be fine-grained.

(3) Different Sets of Ramsey Sentences

Human psychological states are implicitly defined by sets of Ramsey sentences (Lewis 1972). So are Martian psychological states. These sets of Ramsey sentences will generally be separate. We should not have to appeal to roles in Martian psychology to implicitly define the properties that figure in human psychology. So different psychological terms will figure in each set.

Human psychological states are functionally defined in terms of inputs, outputs, and beliefs, desires, intentions and other psychological states (Psy1, Psy2, ...):

$$\begin{aligned} \text{bel } p &= f(\text{bel } p', \text{des } q, \text{intention } x, \text{Psy1}, \dots, \text{stimulus1}, \dots, \text{behaviour1}, \dots) \\ \text{des } q &= f'(\text{bel } p, \text{des } q', \text{intention } y, \text{Psy2}, \dots, \text{stimulus2}, \dots, \text{behaviour2}, \dots) \\ &\dots \end{aligned}$$

We then follow the Ramsey-Lewis method to give a functional characterisation of all human psychological types at once in terms of their interrelations.² Functionalism says the same process is available to characterise Martian psychology, but functionalism need not be committed to Martians having the same collection of interrelated mental state types. So the specifications that enter into a functionalist treatment of Martian mental states (MPsy1, MPsy2, ...) need not mention any human psychological categories:

$$\begin{aligned} \text{MPsy1} &= g(\text{MPsy2}, \text{MPsy3}, \dots, \text{stimulus } 1, \dots, \text{behaviour1}, \dots) \\ \text{MPsy2} &= g'(\text{MPsy1}, \text{MPsy3}, \dots, \text{stimulus2}, \dots, \text{behaviour2}, \dots) \\ &\dots \end{aligned}$$

² Wadham (2016) points out that belief may be a cluster or prototype concept, based on having a sufficient number of relevant properties. Then the definition will be more complex, where some conditions are not individually necessary but certain conjunctions of them are.

We should distinguish between two different kinds of Martian. Martians* are very like humans in the fine-grained functional organisation of their psychological processing, but that processing happens to take place in a radically different substrate – silicon perhaps. Here the intuition is that such creatures would operate in all relevant respects, both internally and externally, just like humans (although perhaps on a different length- or timescale). An intuition of substrate-neutrality underpins the conclusion that Martians* have beliefs, desires, perceptions, working memory, and all the rest.

The second anti-parochial functionalist intuition discussed above motivates a second kind of case, Martians**. Martians** have a quite different psychological set up. Here a difference in substrate is irrelevant. The functional organisation of Martians** is unlike anything humans have ever known. For example, the way they organise and process information may be radically different. Nevertheless, they seem to act purposively, pursue long-term projects, deal effectively with changing environmental contingencies, build tools and artefacts, communicate with each other, work together in complex social organisations, and so on. Their psychology is so alien that we humans would never be able to empathise with them or see the world as they do. Still they do seem to have a psychology, to behave on the basis of inner psychological states. The octopus may be a real-life example of our hypothetical Martian**. It shares our organic substrate, but our last common ancestor with the octopus was some relatively simple probably worm-like organism that almost certainly did not have beliefs (Godfrey-Smith 2017). Yet we have the strong sense that octopuses are intelligent, with some kind of interesting if very different form of psychological organisation; that is, that they have psychological states.

We can distinguish two potential intuitions about Martians**: that they have psychological states and that they have beliefs. Granted that we set up the case so that Martians** seem to have a psychology of some kind, is it also plausible that they have beliefs? It may seem unlikely that a psychofunctionalist specification of what it is to believe that *p* would extend to Martians**, but what about commonsense functionalism about belief? Commonsense functionalism was designed with anti-parochialist motivations in mind (Braddon-Mitchell and Jackson 2007) and this is the form of functionalism relied on by Clark in the extended mind debate (Clark 2008, p. 88, 240). But even commonsense functionalism appeals to interrelations between various mental states, for example between belief, desire (as opposed to drives or

hedonic states) and intention. There is no reason why this interrelated collection of mental states should mark the bounds of the psychological.

So the more secure anti-parochialist motivation for functionalism implies that Martians** have psychological states, not that they have beliefs, desires or any other human mental state. Martians** do not then imply that the functional specification of beliefs must be drawn broadly. They show only that any functionalist characterisation of what it is to be a psychological state must be sufficiently broad to include the set of fine-grained relations amongst Martian psychological states exhibited by Martians**, as well as the (very different) set of fine-grained relations amongst human psychological states exhibited by humans.

The distinction between psychofunctionalism and commonsense functionalism need not align with the distinction between fine-grained and coarse-grained functional specifications. A Ramsey sentence that tied belief to just a single non-commonsense psychological category would be psychofunctionalist but coarse-grained. A Ramsey sentence that posited complex interrelations between beliefs, desires, intentions, perceptual states, emotions, sensations, moods, memories and sentence meaning, for example, would be commonsense and fine-grained. Our question is about fineness of grain.

We saw above that both Clark and Sprevak rely on the idea that a motivation for functionalism – that Martians could have mental states – suggests that functional specifications will be sufficiently coarse-grained so as to apply to some actual human informational resources outside the skin. Consider the well-known case of Otto, who uses entries in a notebook as a way of remembering addresses. Sprevak envisages a Martian with a storage and recall system inside its head that works like Otto's notebook (2009, p. 508). He argues that functionalism entails that the notebook is cognitive in both the Martian and the human (Otto) case. However, the functionalist intuition that the Martian is a cognitive agent could at most entail only that the notebook falls under one of the categories of Martian psychology, e.g. MPsy1. A state in the Martian's internal notebook of functional type MPsy1 could stand in the right relations to the other states of Martian psychology to count as a psychological state of some kind. But it does not stand in the right functional relations with human beliefs, desires, intentions and so on to count as a human belief, since it does not stand in any causal relations with human psychological states. A fortiori, it does not stand in the right functional relations with human psychological properties to qualify as any other kind of human psychological state.

Now consider the notebook of the human, Otto. Nothing in the case speaks against human beliefs being individuated by relatively fine-grained functional roles, in which case the notebook does not stand in the right relations to human psychological states to satisfy the functional definition of any of them (the functions f, f', \dots above). That would be so whether or not the notebook was internal or external to Otto's skin. Nor does it stand in the right relations with MPsy2, MPsy3, etc., to count as a Martian cognitive state (the functions g, g', \dots above), since it is not, by hypothesis, causally connected to any states of Martian psychology at all. So functionalist treatments of particular kinds of human psychological state (belief, desire, etc.) do not licence the conclusion that the notebook contains beliefs; nor do functionalist treatments of particular Martian mental states licence the conclusion that the (human) Otto's notebook contains any Martian-type mental state.

In short, the possibility of a Martian** psychology in which interactions with a notebook were integrated with other Martian** psychological states in such a way that states of the notebook count as states of Martian** psychology would not immediately have the consequence that states of (human) Otto's notebook are psychological states of any kind. Both Otto and the Martian** have the right kind of functional organisation to fall under the very general kind, *psychological system*. Otto's psychology is human psychology, which may well individuate beliefs in a sufficiently fine-grained way that states of the notebook do not count as beliefs (e.g. because they are not sufficiently fluidly integrated with the rest of Otto's beliefs). Martian** psychology has a different inventory of psychological states, by reference to which the Martian**'s notebook counts as an instance of state MPsy1**, say. That does not imply that human Otto's notebook states are instances of MPsy1**. They are not causally connected to the other MPsy** states by which MPsy1** is individuated.

Whether things fall out that way depends upon three issues: how human psychological states are to be individuated; how Martian** psychological states are to be individuated; and what it takes to fall under the general category of being a psychological system. Those issues are where the hard work needs to be done to decide if there are any actual examples of cognitive extension. In particular, as is widely recognised, the answer will depend on how coarse- or fine-grainedly human psychological states ought to be individuated. My aim here is not to argue against the extended mind hypothesis, but to resist the argument that a major motivation for functionalism entails radical cognitive extension. That argument does not go through once we

recognise that the states in a psychological system are functionally defined by reference to causal relations to other states drawn from the same kind of psychological system. Intuitions about the psychological states of organisms with a very different functional organisation from ours – with a very different psychology – therefore get no grip on external resources with which humans interact.

We could instead consider a Martian that matched the fine-grain psychology of human belief, desire, etc., but who also had a notepad as an extra internal resource, interacting with beliefs, desires, etc. in the same ways, functionally, that Otto's external notepad interacts with his beliefs, desires, etc. If the functionalist intuition were to require us to count the Martian's internal notepad as cognitive, that would indeed entail that Otto's external notepad is cognitive. But does functionalism entail that the Martian's internal notebook is cognitive? The fact that it is located inside the Martian's head cannot be determinative, on a functionalist treatment (Wheeler 2010a). Treating him as a Martian*, we conclude (rightly) from the similarity in his internal organisation that he has beliefs, desires, etc. But that says nothing about whether his notebook is a resource that meets any of these fine-grained specifications of human psychological states.

To answer that question we need an account of what it takes to be a belief (where belief is one of the psychological state types instantiated by humans, and possibly other species). Again, that question is where the action takes place. Notice that it will not be enough to have an independently-motivated locationally-uncommitted account of what it takes for a state to be cognitive in general, as some have argued (Walter 2010, Wheeler 2010b). That will just tell us what it is for an arbitrary collection of interrelated functionally-defined states to count as cognitive – a test that applies to a whole organism and its full inventory of states. That states of a notebook could count as psychological within some functional organisation or other does not yet show that they count as psychological when causally related to the categories of human psychology (belief, desire, intention, etc.). That still depends on how coarse-grained or fine-grained the right functionalist characterisation of beliefs should be.

A final question is whether, even if they are not beliefs, states of Otto's notebook fall under some other psychological state type. Maybe the advent of writing brings with it the appearance of new psychological categories. The written artefacts people interact with in the right way would then count as psychological, in the spirit of the extended mind intuition,

without falling under any of the psychological categories applicable to pre-literate humans (belief, desire, intention, etc.). Whether that is so depends on what it takes to be a psychological state, which I have not attempted to answer here. But notice that the conclusion would be less radical than the standard extended-mind claim that Otto has beliefs. We already rely on categories like *diary*, *portable written record* and *electronic address book* to explain patterns in human behaviour. Treating these systematically, a category like *readily-accessible artefact-based memory* could turn out to qualify as psychological, taking its place alongside semantic memory and episodic memory in explanations of human behaviour.³ This possibility does not entail that the category of *belief* must be functionally individuated in a coarse-grained way.

(4) Conclusion

So we return to the question which is widely-recognised to lie at the heart of the extended mind debate: whether there are extra conditions on a state's being a belief which limit the ambit of cognitive extension (Clark & Chalmers 1998; Adams & Aizawa 2001, 2008), foreclosing the problematic consequence that every informational resource with which a human interacts is part of her extended mind (Rupert 2004, 2009; Sprevak 2009). That debate cannot be resolved by a general argument that the motivations for functionalism require coarse-grained functional specifications of categories like belief, pace Clark and Sprevak. Functionalism remains a viable ontology of mental states, and the fate of 'extended functionalism' (Clark 2008) turns on whether external information resources have the kinds of connections to the rest of human psychology that would, whether they were located inside or outside the skin, make them count as a belief, a desire, an intention, or any other particular type of human psychological state.

The fineness-of-grain question is important for functionalism in general, not just for those interested in the hypothesis of extended cognition. The anti-parochialist motivation for functionalism would seem, at first glance, to motivate a relatively coarse-grained way of individuating psychological states. But that does not follow. We need to differentiate between functionalism about what it takes, in general, to have a collection of psychological states that amount to a psychological system, and functionalism about what it takes to be a belief. A fine-grained way of individuating beliefs is compatible with a functionalism that counts organisms

³ There is an issue about whether these would be psychological states of Otto or of the hybrid system consisting of Otto-plus-notebook (Miyazono 2017).

with quite a different functional organisation, and without beliefs, as also being psychological systems, and hence as having psychological states of their own kind.

Nicholas Shea

nicholas.shea@sas.ac.uk

November 2018

Acknowledgements

The author would like to thank Ellen Fridland, Mark Sprevak and two anonymous referees for helpful comments and criticism. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 681422 (MetCogCon).

References

- Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14, 43-64.
- Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Malden, MA: Blackwell.
- Braddon-Mitchell, D., & Jackson, F. (2007). *The philosophy of mind and cognition* (2nd ed.). Oxford: Blackwell.
- Chalmers, D. (2008). Forward to Clark (2008).
- Clark, A. (2007). Curing Cognitive Hiccups: a defense of the extended mind. *Journal of Philosophy*, 54(4), 163-192.
- Clark, A. (2008). *Supersizing the Mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7-19.
- Godfrey-Smith, P. (2016). *Other Minds: The Octopus, the sea, and the deep origins of consciousness*. New York: Farrar, Straus and Giroux.
- Godfrey-Smith, P. (2017). The evolution of consciousness in phylogenetic context. In K. Andrews & J. Beck (eds.), *Routledge Handbook of Philosophy of Animal Minds*. New York: Routledge, pp. 216-226.
- Lewis, D. (1972). Psychophysical and Theoretical Identification. *Australasian Journal of Philosophy*, 50, 249-258.
- Miyazono, K. (2017). Does functionalism entail extended mind? *Synthese*, 194, 3523–3541.

- Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, 101, 389-428.
- Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.
- Sprevak, M. (2009). Functionalism and extended cognition. *Journal of Philosophy*, 106, 503-527.
- Wadham, J. (2016). Common-sense functionalism and the extended mind. *Philosophical Quarterly*, 66(262), 136-151.
- Walter, S. (2010). Cognitive extension: the parity argument, functionalism, and the mark of the cognitive. *Synthese*, 177, 285-300.
- Wheeler, M. (2010a) Minds, things, and materiality. In L. Malafouris & C. Renfrew (eds.), *The cognitive life of things: Recasting the boundaries of the mind*. Cambridge: McDonald Institute Monographs.
- Wheeler, M. (2010b). In defense of extended functionalism. In R. Menary (ed.), *The Extended Mind*, pp. 245-270. Cambridge, MA: MIT Press.