Open Research Online

[Book Review] "The Metaphysics of Representation" by J. Robert G. Williams

Journal Item

Version: Accepted Manuscript

The Metaphysics of Representation. By J. ROBERT G. WILLIAMS. (Oxford: Oxford University

Press, 2019. Pp. xxvi + 214. Price £50.00.)

In *The Metaphysics of Representation*, J. Robert G. Williams aims to develop a reductive

metaphysics of representation—that is, to explain how representation arises in a fundamentally

non-representational world. This is a difficult task, to put it mildly, and previous attempts have

generally resulted in either handwaving or extreme complication. Happily, Williams finds a

middle ground. On the one hand, he provides enough detail to convince the reader that the

account might actually work. On the other hand, while the account has many moving parts,

Williams draws such clear and intuitive conceptual distinctions between those parts that the

reader never feels lost. Williams' organisation of the material is second only to the material

itself.

As I understand him, Williams is most interested in mental and linguistic representation.

His strategy, however, is not to reduce mental and linguistic representation to non-

representational facts directly. Rather, the strategy is this: linguistic representation is to be

reduced to mental representation; mental representation is to be reduced to a set of low-level

perceptual and action-guiding representations; and those low-level representations are to be

reduced to the non-representational world.

This strategy of divide and conquer is powerful: it allows Williams to give a *different*

account of representation at each level. And, in particular, it allows him to adapt and combine

extant accounts of the metaphysics of representation, putting each account to work at the level

at which it is most plausible. Linguistic representation, following a broadly Lewisian approach, is

to be grounded in conventions of truthfulness and trust—where a convention differs from a

mere regularity in part in virtue of participants in the convention having certain beliefs about

the regularity. Mental representation, in turn, is grounded by radical interpretation. Here,

Williams departs from work in the area by Quine, Davidson and Lewis, by allowing low-level representations to feed into the interpretative process. And, finally, the perceptual and action-guiding representations that feed into radical interpretation are to be grounded teleosemantically, reduced to functions and causes. Here, Williams adopts (and extends) recent work by Neander.

The centrepiece of the account is the middle layer: the use of radical interpretation to ground mental representation. This is where Williams makes the most decisive advances. The view is roughly this: when *x* has a belief, desire and/or concept with a particular content, the belief, desire and/or concept has that content because the best rationalisation of *x*'s dispositions to act (read: action-guiding mental states), given *x*'s courses of experience (read: perceptual states), assigns that content to the belief, desire and/or concept. *Rationalisation*, here, is given a substantive reading, understood in terms of what is both *epistemically* and *morally* optimal: 'a substantively rational agent believes as they ought (or as is permissible), given their evidence, [and] acts as they ought (or as is permissible), given their options and beliefs' (p. 26).

In a three-chapter highlight of the book, Williams uses radical interpretation to derive substantive local theories of reference for a variety of concepts. Williams is careful to explicitly draw out the assumptions of these derivations, so that the results are effectively conditionals. For example, roughly (pp. 40–42): *if* Sally's thinking has a language-like structure, *if* her reasoning processes has a natural-deduction-like form, *if* we can identify the rules that govern that process independently of content-determination, and *if* Sally has a concept *c* for which she follows the rules:

From A*,* B *derive* AcB

From AcB *derive* A

From AcB *derive* B,

then, assuming a substantively rational agent would follow those rules just in case *c* denoted conjunction, radical interpretation implies that *c* denotes conjunction. Now, that's a pretty bold antecedent just to assign conjunction to a concept, but it illustrates the basic idea. And Williams goes on to derive more substantive conclusions. For example, given comparable assumptions, Williams goes on to argue that unrestricted universal quantification is possible, that 'explanatory concepts' denote whatever natural property is in the vicinity (so that natural properties are 'reference magnets'), that the concept of wrongness is referentially stable, and that contents are shared widely across populations. These are impressive, concrete results to derive from a general metaphysics of representation, even given bold assumptions.

Williams makes several novel and important moves in developing this account. One is the appeal to representational states; another is the appeal to 'substantive rationality'; a third is the attempt to derive concrete results. Most importantly, perhaps, is the strategic realisation that radical interpretation is not a competitor of 'local' theories of reference. Given relevant assumptions, for example, radical interpretation *explains why* inferentialism appears to be true of logical connectives, *why* reference magnetism appears to be true of 'explanatory concepts', *why* the concept of wrongness is referentially stable, and so on. Within Williams' framework, we have a principled, theoretical reason to expect different local theories of reference to hold of different concept groups: this patchwork of local theories is predicted by a unified, foundational metaphysics of representation.

For me, however, an important question remains. As noted above, Williams derives the local results by making substantive psychological assumptions. For creatures whose cognitive processes run on a broadly Fodorian language-of-thought, Williams' account looks compelling. But we are probably not such creatures. And, if that is right, there is little principled reason to expect Williams' local results—about logical connectives, explanatory concepts, etc.—to apply to *us*. This raises the following question: Why do Williams' local results appear to fit *our*

3

concepts so well? Why do assumptions that (probably) don't hold of *us* allow Williams to recover what many take to be the most plausible, extant local theories of reference for *our* concepts? Williams does not raise this question, but I suggest that it deserves an answer.

An optimistic answer is this: when it comes to radical interpretation, the specific psychological make-up of a creature is not that important. So, while it was helpful for Williams to derive local results for someone with an idealised psychology, the results themselves transfer over to us. This would be a nice answer, but it's one that needs arguing for.

A 'philosophy first' answer is this: the fit of Williams' result with our concepts is evidence that we are creatures whose cognitive processes run on a broadly Fodorian language-of-thought. Caution is required here, however, as general metaphysical results do not obviously carry significant weight when it comes to substantive questions about cognitive architecture.

A pessimistic answer is this: it is a mistake to think that radical interpretation can yield substantive local results. Rather, Williams has provided just-so stories to justify whatever local theories of reference he happens to find plausible. Others, in turn, could 'derive' *their* preferred local theories from the same psychological assumptions, by focusing on different aspects of our cognitive processes and by adjusting the assumptions about substantive rationality. This pessimistic answer would undermine the promise of Williams' view, and I hope it can be ruled out. But we cannot assume that it will be.

To ascertain the full implications of Williams' work, then, such questions need answers. Whatever those answers turn out to be, however, this is an excellent book, and highly recommended. Williams extends extant work in insightful and important ways, combining it into a powerful and cohesive account. *The Metaphysics of Representation* is essential reading for anyone whose research engages with metasemantics, and is recommended for anyone interested more generally in linguistic, conceptual or other forms of representation.

MARK PINDER

The Open University