

Review of *Representation Reconsidered*

Mark Sprevak
University of Edinburgh

11 May 2010

WILLIAM M. RAMSEY, *Representation Reconsidered*. Cambridge: Cambridge University Press, 2007, 268 pp., £54.00 (hardback). ISBN 978-0-521-85987-5.

William Ramsey's *Representation Reconsidered* is a superb insightful analysis of the notion of mental representation in cognitive science. The book presents an original argument for a bold conclusion: partial eliminativism about mental representation in scientific psychology. According to Ramsey, once we examine the conditions that need to be satisfied for something to qualify as a representation, we can see those conditions are not fulfilled by the 'representations' posited by much of modern psychology. Cognitive science—or at least large swathes of it—has no warrant for positing representations.

The structure of Ramsey's argument repeats a familiar eliminativist strategy (Churchland 1981; Mallon et al. 2009; Stich 1983). First step: argue that in order for something to be an *X*, it must satisfy a certain description *D* (say, beliefs must satisfy the description given in folk psychology). Second step: argue that to the best of our knowledge, nothing satisfies description *D* (e.g. folk psychology is false). Third step: conclude that since nothing satisfies description *D*, there are no *X*s (no beliefs). Here is how the strategy is played out in the book. First, Ramsey argues for certain minimal conditions that a representation must satisfy (what he calls the 'job description'). Second (this takes the bulk of the book), he considers the ways in which our best psychological theories use the notion of representation. Ramsey argues that none of these uses satisfy the job description associated with representation. (A wrinkle is that some representations—those posited by the classical computational theory of cognition—*do* qualify as genuine representations. But, Ramsey claims,

classical theories are in a minority in cognitive science, and their hold on the field is shrinking.) Therefore, Ramsey concludes, at least in most of cognitive science, there are no mental representations.

That much is the negative component of Ramsey's argument. The positive component is that the internal states and mechanisms currently labelled 'representations' can be re-glossed as causal relays or dispositions. Replacing representations with relays or dispositions will, according to Ramsey, have little impact on the practice of cognitive science. Empirical cognitive science can continue in roughly the same way minus talk of representational states. In a nutshell, Ramsey's claim is that representation talk adds nothing of explanatory value to (most of) our psychological theories, and is therefore unwarranted. Mental representations are wheels that do not turn anything in psychological explanations.

Let us consider each step of Ramsey's argument.

First, step one: the job description challenge. This involves coming up with a description (*D* above) of the essential or core features of representation, which we can use in assessing whether scientific theories that appear to posit representations really do or not. Ramsey highlights two conditions for something to be a mental representation: *non-derived intentionality* and *causality*. For a state to be a mental representation it must: (i) be capable of having original intentional content, and (ii) interact causally with other cognitive states. To these conditions, Ramsey adds a third (and this does most of the work for him). He claims that (i) and (ii) need to be somehow linked: the causal role that a representation plays should be determined by its intentional content. For something to be a representation it must *function as such*: its intentional content should be relevant to either how the system works, or our best explanation of how the system works. If the intentional content is just along for the ride, and does no causal or explanatory work, then there seems no reason to assume that the state in question is specifically *representational*, rather than, say, a causal relay with the same effects.

Second, step two: arguing that candidate uses of 'representation' in cognitive science do not meet the job description. First, Ramsey focuses on the exception: the classical computational theory of the cognition (CCTC). According to Ramsey, CCTC is unique in cognitive science in that it is committed to representations. CCTC's commitment arises in two ways, which Ramsey calls *IO-representations* and *S-representations*. This labelling is somewhat confusing, since the names designate different *ways* in which the CCTC is committed to representations, not different *types* or theories of representation. It is an open question whether a single type of representation satisfies both roles.

First, commitment via the IO-representation-role. Ramsey is not an eliminativist about non-mental representation and he observes that part of what cognitive agents

do is transform non-mental representations (e.g. heard linguistic input into spoken linguistic output). Representations are therefore needed as the gross inputs and outputs of cognitive agents. Cognitive science aims to explain how this transformation works. CCTC discharges this obligation by positing computations, which by their nature take representations as input and yield representations as output. Computations are also, by their nature, compositional: they are made up from parts, steps, which are often computations in their own right. These mini-computations require their own representations as inputs and outputs. Therefore, CCTC is committed, not just to representations as the gross inputs and outputs of a cognitive agent, but also to internal representations.

Second, commitment via the S-representation-role. Ramsey's idea is that CCTC theories often explain by attributing models of the world to the agent. These models tend to encode structural features of the world (e.g. spatial relations between objects) by their own structure. By reasoning about the structure of the internal model, cognitive agents can make inferences about the structure of the world. Ramsey argues that we can only understand the success of these agents—how these agents manage to make successful inferences *about* the world—if we understand the elements of their models as *representing* features of the world. Otherwise, it would be mysterious why examining facts about the structure of their internal models would allow agents to succeed. Hence, CCTC is committed to positing internal representations.

What detracts from Ramsey's argument is that he goes on to infer a less plausible conclusion about how the *content* of CCTC representations gets fixed, which is conflated with the above claim about commitment. Ramsey claims that an internal model has its content in virtue of an *isomorphism* between its structure and the structure of the intended target system. There are two points to make about this. First, it is not clear how it helps Ramsey's argument, it confuses a representational commitment (*viz.* explanation of inference requires internal models) with a positive proposal for a theory of content. Second, isomorphism-based theories of content face fearsome objections: isomorphism is a symmetrical relation, but representation is not; isomorphism is too weak a condition for representation, it entails massive indeterminacy in content (e.g. there is an isomorphism between the structure of letters in my car registration and the structure of letters of my dog's name, but they do not represent each other; isomorphism is also too strong a condition for representation, a representation need not be isomorphic to what it represents (e.g. a crude map of the UK can represent the UK, even if it entirely fails to be isomorphic to the UK). Ramsey considers some of these objections, but he does not succeed in defanging them. Moreover, it is unclear why he needs to take on the task of defending this, or indeed any, theory of content. It is not clear how the observation that CCTC posits structured models that function as surrogates for the world in reasoning entails, or indeed lends any support, to the conclusion that the representation

relation obtains *because* of this structural similarity. Just because I use a model that shares structure with what it represents, that does not mean that the model represents because of its shared structure. Cognitive agents may infer by examining the structure of mental models that function as surrogates for the world, but the question of how those models get their particular representational content is entirely separate (it may be causal, historical, etc.).

We will return this point shortly, but let us focus on the eliminativist part of Ramsey's argument. Ramsey moves from the CCTC to cognitive neuroscience and connectionism. Both theories employ what he calls the *receptor notion* of representation: *X* represents *Y* if the occurrence of *X* is reliably caused by, or is nomically dependent on, the occurrence of *Y*. At least on the face of it, the receptor notion is not particularly representational: nomic dependence between two conditions is common, but we are rarely tempted to attribute representations (think, thunder and lightning). Ramsey introduces Dretske (1988)'s account, which tries to capture what makes some nomic dependencies distinctively representational. On Dretske's view, a state *X* in a system represents *Y* if (i) *X* stands in the appropriate nomic dependency relation to *Y*, and (ii) *X* became incorporated into the system's processing *because* of that dependency, in other words, *X* serves the *function* of selectively responding to *Y*.

Ramsey claims that even with Dretske's appeal to function, the receptor notion is still too weak. A firing pin in a gun reliably mediates between the pulling of the trigger and the discharge of the round. The firing pin is built into the gun for precisely this reason. Yet we would not say that the firing pin thereby qualifies as a representation. Similarly, a spark plug mediates between an accelerator pedal and a drive shaft, but we would not say that the spark plug represents the accelerator pedal. Talk of representation in these cases is overblown: it does not add explanatory value to saying merely that the state is a reliable causal relay. Ramsey's objection to the receptor notion is that violates his third condition in the job description of a representation: representational content should do some explanatory work. The problem with the receptor notion is that the nomic dependency relations do all the explanatory work. The *content* does not have an explanatory role over and above the effects involved in the nomic correlation. This goes for firing pins and spark plugs and just as much for 'edge detector' cells in the V₁ cortex. Perhaps content has an explanatory role because, on Dretske's view, the content simply *is* or *is determined by* the nomic correlation? No, as the firing pin and spark plug cases show, nomic correlation is not sufficient for representational content, and so cannot be it or determine it. Ramsey concludes that employment of the receptor notion should be understood as involving a commitment to no more than the non-representational causal relations picked out by the nomic dependency relation. We should be eliminativist about talk of representation in cognitive neuroscience,

connectionist modelling, or any theory that relies on the receptor notion.

That, in abbreviated form, is Ramsey's argument. How successful is it?

On the positive side, I think that Ramsey presents a powerful challenge to naturalistic theories of representation. He shows that a state can satisfy Dretske's conditions without being a representation. On the negative side, I do not see how Ramsey's eliminativist argument succeeds. The problem stems from Ramsey's own exception, the CCTC. As observed above, Ramsey assumes (i) the S-representation-role is the sole property of the CCTC, and (ii) only an isomorphism-based account of content fulfills that role. Both assumptions appear to be wrong, and this causes trouble for Ramsey's eliminativism.

First, the commitment to representation engendered by the S-representation-role is not unique to the CCTC. To recap, the S-representation-role is that cognitive agents make inferences about the world by reasoning about internal models, and we best explain the success of those agents if we assume that those internal models represent the world. But it is hard to see anything distinctively CCTC about this explanatory move. It is commonplace in cognitive neuroscience, connectionism, indeed, all areas of cognitive science, to explain behavioural success in terms of the agent's inferences about internal models. Different theories posit different models, but they alike apply the strategy of explaining worldly success in terms of the agent's inferences about those models. For example, in order to explain the success of an agent in making inferences about *edges* in her visual field, one typically assumes that elements in her internal model (e.g. the activity in her V1 cells) *represent* distal edges. This explains why the agent's inferences about that activity contribute to her worldly success. It is hard to see why Ramsey thinks CCTC is exceptional; if CCTC has the S-representation-role commitment, then so do these other theories.

Ramsey's second assumption also causes trouble. As noted above, how our mental models get their content need not be answered by an isomorphism-based theory. Ramsey himself is explicit that he does not challenge the receptor notion as a theory of content, only as a theory of what makes something a representation. This leaves open that what satisfies the receptor notion, by itself, may not fulfill the job description of a representation, but the wider explanatory role that it plays in explaining successful behaviour may justify its labelling as a representation rather than a causal relay. Consequently, as far as Ramsey's argument goes, connectionist models and cognitive neuroscience may be *right* to label internal states that satisfy the receptor notion *representations*, provided such a labelling does explanatory work in that it explains the success of the agent in reasoning about the world—which like CCTC models, it generally appears to do. So the receptor notion can gain Ramsey's truly representational credentials via its wider explanatory role in explaining the success of the agent, just as his isomorphism-based representations do. The mistake,

which strangely enough is one that Ramsey repeatedly points out, is to think that the receptor notion, by itself, provides a reductive theory of *representation*, whereas it only provides one component, a theory of *content*.

Ramsey argues that what is special about isomorphism-based representations is that they stand in for something, they are surrogates for reasoning. But it is far from clear why a receptor based notion cannot fulfill the role of a surrogate too. And the receptors posited in cognitive neuroscience typically do: a face detector or edge detector are understood, not just as detectors in isolation, but as part of our wider model of the world, and their activity is considered an apt surrogate for the presence of distal *faces* or *edges* in reasoning. Again, the point is that the distinctive S-representation-role that Ramsey claims is unique to the CCTC and uniquely filled by an isomorphism-based theory of content is not unique to classical architectures and can be filled by other theories of content (causal, historical, etc.).

This just scratches the surface of Ramsey's excellent book. Whether the partial eliminativism succeeds or not, the overriding virtue of his book is that it frames, and attempts to answer, important questions: What is the job description of a representation? Which of our psychological theories posit entities that fulfill that description? These questions have not received sufficient attention in a literature dominated by the search for theories of content. In his job description challenge, Ramsey makes clear how a theory of content is a distinct project from a theory of representation. I would warmly recommend Ramsey's book to anyone working on representation.

Acknowledgements

I would like to thank Jim Bogen, Lisa Damm, Natalie Gold, Edouard Machery, and all the participants in the Pittsburgh reading group on cognitive science, as well as William Ramsey and Oron Shagrir for comments on this review.

References

- Churchland, P. M. 1981. 'Eliminative materialism and the propositional attitudes'. *The Journal of Philosophy* 78:67–90.
- Dretske, F. 1988. *Explaining Behavior*. Cambridge, MA: MIT Press.
- Mallon, R., E. Machery, S. Nichols and S. P. Stich. 2009. 'Against arguments from reference'. *Philosophy and Phenomenological Research* 79:332–356.

Stich, S. P. 1983. *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.