Synthese (2017) 194:4197–4214
DOI 10.1007/s11229-014-0644-6

Must cognition be representational?

William Ramsey

Received: 13 March 2014 / Accepted: 19 December 2014 / Published online: 21 January 2015

Abstract In various contexts and for various reasons, writers often define cognitive processes and architectures as those involving representational states and structures. Similarly, cognitive theories are also often delineated as those that invoke representations. In this paper, I present several reasons for rejecting this way of demarcating the cognitive. Some of the reasons against defining cognition in representational terms are that doing so needlessly restricts our theorizing, it undermines the empirical status of the representational theory of mind, and it encourages wildly deflationary and explanatorily vacuous conceptions of representation. After criticizing this outlook, I sketch alternative ways we might try to capture what is distinctive about cognition and cognitive theorizing.

Keywords Cognition · Demarcation criteria · Marr's levels · Representationalism · Representation demarcation thesis · Folk psychology

1 Introduction

What kinds of states and processes should we characterize as cognitive in nature? A related question is this: what sorts of theories about brain processes merit the label of cognitive, as opposed to, say, purely physiological? One traditional answer to this question is that cognition is a process that in some way or other involves inner content-bearing, representational states and processes, and thus cognitive theories are those that are about those representational states and processes. According to this perspective, the link between representationalism and cognitive science is quite strong, and by that

W. Ramsey (✉)
Department of Philosophy, University of Nevada, Las Vegas, NV, USA
e-mail: wramsey@unlv.edu

I mean it is treated as conceptual in nature. It is part of how many folks define the discipline of cognitive science itself.

Here I will argue that tying cognitive science to representationalism in this way is a bad idea. We should not make the cognitive status of states and processes in any way dependent upon the presence of representations, and we should not make the cognitive status of any theory dependent upon whether or not it posits representations. My focus is not on the status of representations as explanatory posits, as such; instead, it is on their role as a demarcation criterion—on their value as a "mark of the cognitive". Even if you think cognitive scientists must invoke representations to explain a wide array of cognitive capacities and processes, I'll argue you should nevertheless reject the use of representations to define cognitive processes and theories.

To show this, my essay will have the following format. First, I'll say a bit more about the representational demarcation criterion and where I think it comes from. Next, I'll offer three arguments against using representations to demarcate cognitive processes and explanations. Roughly, my arguments will be that doing so puts an unnecessary restriction on our theorizing, it undermines our understanding of the representational theory of mind, and it encourages bad theory development. After presenting these arguments, I'll make some suggestions about how I think we *should* think about cognitive activity and cognitive science. I'll suggest that the demarcation problem is not as pressing as some may think; moreover, if we are going to try to define cognitive science, we should do in terms of what we are trying to explain, not in terms of the nature of the explanation.

Before getting started, however, I want to emphasize one crucial point of clarification. In recent years, there has been growing skepticism regarding the explanatory value of representational posits in cognitive science, and many writers have presented compelling arguments against thinking the mind even uses representational states and structures (see, for example, van Gelder 1995; Chemero 2011; Hutto and Myin 2012). I have, to some degree, contributed to this skepticism with my own arguments for the thesis that many things described as representations in cognitive theories are actually not representations, given their assigned functionality (Ramsey 2007). Of course, any skepticism about the explanatory role of representations in cognitive science will, by default, provide indirect support for my thesis that cognitive science should not be demarcated by appealing to representations. But it is important to be clear that my aim is not to repeat these anti-representational arguments, nor is it to promote skepticism about the existence of representations in the brain. Instead, I want to argue that even if you are strongly committed to representationalism, you should nevertheless avoid conceiving of cognition in representational terms. In other words, even if you find the arguments against representationalism unconvincing or reject the new anti-representational perspectives associated with radical embodied cognition or dynamic systems theory, you should still avoid defining cognition by appealing to representations. My goal here is to establish not an anti-representational thesis, but rather an anti-representation-as-definer-of-cognition thesis.

2 Background: the representation demarcation thesis (RDT)

It is hardly news that traditional, mainstream cognitive science is closely tied to representationalism and that a basic assumption of many cognitive psychologists is that we should explain cognitive processes by appealing to information-carrying states and operations. But often times this outlook is taken further, and the positing of representations is treated as more than just a common theoretical practice; it is also viewed as defining or demarcating the discipline itself. Cognition is defined as those operations and processes that involve inner representations, and hence cognitive theories are those that describe those inner representational processes. For example, Fodor and Pylyshyn, tell us that "it's the architecture of representational states and processes that discussions of *cognitive architecture* are about. Put differently, the architecture of the cognitive system consists of the set of basic operations, resources, functions, principles, etc…whose domain and range are the *representational states* of the organism" (1988, p. 10). Similarly, Bermudez tells us "The idea of cognition without representations or computation is almost a contradiction in terms" (Bermudez 2010, p. 415). Thagard tells us that, "The central hypothesis of cognitive science is that thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures. While there is much disagreement about the nature of the representations and computations that constitute thinking, the central hypothesis is general enough to encompass the current range of thinking in cognitive science…" (Thagard 2014). And finally, Rowlands provides us with this definition: "A process P is a cognitive process if [and only if]: (1) P involves information processing—the manipulation and transformation of information-bearing structures…(3) This information is made available by way of the production, in the subject of P, of a representational state…" (Rowlands 2010, pp. 110–111).

This is just a small sampling of the sort of claims writers have made that express a strong conceptual link between cognition and representations. Of course, there are many who do not define cognitive processes in this way. Still, the idea that cognitive processes and cognitive theorizing should be categorized by appealing to representations is perhaps one of the most popular foundational assumptions of cognitive science. According to this view, if there is some neurological or even some computational process that does not involve inner representations, then such a process does not really qualify as a form of cognition. Similarly, if there is a theoretical model or framework that does not invoke representations, then, by definition, it is not really a theory about cognition. I will refer to these two closely related claims as the "representational demarcation thesis" (RDT).

The representational demarcation thesis comes in many forms and is generated by various different considerations. For example, it arises from the core assumption that cognitive systems should be viewed as information processing systems. When we ask what 'information processing' means, we are typically informed that it means the invoking, manipulation, storage, and rearranging of inner states and structures that serve to represent various items. The nature of those representations and the sort of operations they are involved in can vary a great deal, from classical computational accounts to connectionist networks. Nonetheless, these radically different models are all commonly described as information processing models because they all appeal to

representations of some form or other. This in turn qualifies them as cognitive models and as legitimate theories of cognition.

RDT is also generated by our understanding of cognitive science as a domain that describes mentality at a level that is more abstract than neuroscience. Marr's (1982) well-known account of explanatory levels posits three explanatory levels, with the top level specifying the particular cognitive task that is to be explained, the middle level specifying the algorithms that do the explaining, and the bottom level describing how all of this is actually implemented in the brain. Marr claimed that cognitive theorizing belongs at the middle level, where he suggested researchers introduce theories about computational operations involving states and structures that are not, strictly speaking, physiological or neurological in nature. Instead, they are often defined as representational states that are invoked and manipulated in the computations. Thus, cognitive science is the science of representational states and the operations that involve them, while neuroscience explains how all of this is physically instantiated.

RDT also arises in various historical accounts of the origin of cognitive science. Cognitive science is what replaced psychological behaviorism. How is cognitive science different? The key difference often invoked is that behaviorist accounts explicitly rejected the explanatory value of inner representations, whereas cognitivists explicitly embrace their explanatory value. Consequently, one popular way in which people characterize the "Cognitive Revolution" is by focusing on the introduction of inner representations to our theorizing about the mind. Representations thereby serve to designate the difference between cognitivism and behaviorism.

The upshot is that our association of cognition with information processing, our understanding of suitable levels of explanation for psychology, and the history of the sciences of the mind all converge to promote RDT—the view that cognitive processes necessarily involve inner representations and cognitive theories must thereby be about the representational states and processes. Yet while RDT is encouraged by these considerations, it is not fully justified by them. Indeed, a deeper analysis of RDT and of what it actually entails suggests that it is a theoretical perspective that is not supported by a proper scientific outlook. The next section offers three reasons for rejecting RDT.

3 Three arguments against RDT

3.1 Argument 1: it puts undue restrictions on psychological theorizing

My first argument against RDT is that it places inappropriate restrictions on our theorizing. If the history of science has taught us anything it is that our theorizing often goes in directions that are unexpected. The final, true theory of a given phenomenon is often much stranger and more iconoclastic than we initially thought possible. Consequently, we have learned that it is unwise to restrict our theorizing by placing artificial boundaries on what an appropriate account is supposed to look like. By "artificial boundaries", I mean restrictions that stem from tradition or convention, or that are grounded in our commonsensical, "folk" conception of the subject. We should not prejudice theory development in favor of so-called received wisdom, nor should we rule out novel and unconventional perspectives that strike us as radical. These are not

deep insights—most of us recognize that the search for truth often requires us to throw out deep-seated assumptions and ideas about a given phenomenon.

A strong case can be made for claiming that equating the representational with the cognitive (or the non-representational with the non-cognitive) restricts our theorizing about the mind in exactly the way we should avoid. In fact, there is reason to think that RDT is grounded in both tradition *and* in our folk conception of the mind. With regard to the former, appealing to inner representations to explain mental phenomena has a very long and glorious history, going back at least as far as Aristotle. While our conception of mental representation has changed greatly over time, its assumed explanatory necessity has been a traditional assumption shared by a great number of thinkers. Notice, for example, the broad array of people who have claimed that intrinsic intentionality—something that presupposes representation—is the "mark of the mental", including writers as diverse as Brentano (1874) and Adams and Aizawa (2001). Within cognitive science, the classical computational theory of mind is widely recognized as the traditional theoretical framework. The classical view treats cognition as a form of symbol-manipulation and thus representations serve as the building blocks of cognition. According to the classical picture, trying to understand the mind without symbolic representations would be like, as Newell suggested, trying to do cellular biology without invoking cells (1980). Indeed, the 'classical physical symbol system hypothesis' was originally put forth by Newell and Simon as the foundational outlook of cognitive science—an outlook that treated symbolic representations as a necessary condition for any form of general intelligence (1976).

At the same time, it seems quite clear that representationalism is a core dimension of our ordinary, common-sense or "folk" psychology. As various writers have noted, central folk posits like the propositional attitudes, images, memories, dreams, certain emotional attitudes and so on are all thoroughly representational in nature. So our tendency to define cognition in representational terms almost certainly stems, at least in part, from this folksy, common-sense conception of what is inside the mind. Even for those who are skeptical about the idea that we explain and predict behavior by using a theoretical framework, such as proponents of the so called "simulation" view who claim we predict behavior by simulating others by using our own cognitive machinery taken "off-line", it is typically admitted that we at least ascribe beliefs, desires, and other mental representations due to our commonsense conception of mentality.

Consequently, there is good reason to think that the demarcation of cognition in representational terms is largely grounded in both very traditional assumptions about how the mind works, and in our folk conception of the mind. This should make us at least a little suspicious that defining cognition this way is scientifically imprudent. There is, in other words, some reason to think that restricting our conception of the cognitive to the representational is a bad idea because such a restriction stems not from robust scientific considerations, but from unscientific prejudices. Of course, that does not mean that all arguments with the conclusion that certain cognitive capacities require representations are bad arguments, or merely unscientific expressions of some bias. I am not claiming that, for example, Chomsky's famous arguments against Skinner that emphasized the essential role of inner representations for linguistic processing were bad arguments. What I *am* suggesting is that there is at least some reason to think our tendency to *define* cognition in representational terms or to simply equate

mental processes with representational processes stems from presuppositions that are not well-founded and that are prone to hinder scientific theorizing. While I am aware of scientifically compelling reasons for thinking that certain areas of cognition involve representations, I am not aware of any scientific reasons for thinking they all must, as RDT implies. An approach that is more consistent with the principles of good science would involve broadening our understanding of cognition to include the possibility of non-representational cognitive states, mechanisms and processes.

But what about those earlier considerations that I claimed helped promote RDT? What about the information-processing outlook and our association of cognition with it? Isn't that a legitimate scientific perspective that commits us to representations? Similarly, doesn't Marr's account of levels commit us to a representational cognitive science for perfectly good scientific reasons, as does the way we distinguish cognitivism from behaviorism?

Here I believe it is important to distinguish a working hypotheses (or starting assumptions) from definitional criteria and necessary conditions. In the case of the information-processing outlook, as I noted earlier, just what exactly this is supposed to be, over and above just a straightforward commitment to some form of representational processing, is not very clear. The processing metaphor is unhelpful, as information is not some sort of substance that can be "processed". If the information processing outlook is simply another way of expressing an assumption that representations are involved in some area of cognition, then its scientific value depends on how far it is taken. If it is functioning as a working hypothesis about various cognitive capacities, then I see nothing wrong as long as it is viewed tentatively with a willingness to abandon in the face of conflicting findings, methodological revisions and promising new alternatives. But if it is really just an expression of RDT, and entails a rigid adherence to representationalism irrespective of future developments, then it is overly confining and ought to be rejected. In other words, it should be possible to put forth accounts of cognitive capacities that aren't, strictly speaking, information processing accounts.

What about Marr's 3 levels of analysis, and the old idea that the middle, algorithmic level is the proper level of analysis for cognitive science? On this view, neurological processes merely implement the higher-level computations or algorithms that are the *real* target of cognitive explanations. If we reject the idea that representations serve to define cognitive processes, then it seems we lose one important justification for understanding the cognitive level of analysis in the way Marr intended. We would, in other words, lose a reason for the anti-reductionist stance that treats neurological processes as only tangentially relevant to understanding cognition.

My reply to this worry is to simply deny there is a problem here. First, there is nothing about Marr's levels of analysis that demands we define the middle level by appealing to representational states and processes. We can still appeal to a higher level of analysis than biology and invoke all sorts of functionally defined mechanisms and states without treating them as playing the role of representations. Non-representationalism needn't entail any sort of extreme reductionism. Second, as a number of writers have pointed out, the traditional image that Marr endorsed of three tidy levels of analysis or explanation is almost certainly naïve and misguided. In cognitive science, there are undoubtedly many more explanatory levels than Marr supposed, and much more overlap between levels as well. Indeed, there are different criteria for sepa-

rating levels, depending on one's explanatory goals and interests. Marr's image of only one level at which cognitive phenomena are to be explained has been abandoned for a much more realistic multi-level picture in which even the lowest levels are also theoretically significant to our understanding of higher-level mechanisms and processes. Third and finally, today relatively few people buy into the old idea that cognitive science needn't really worry about how some psychological process is implemented in the brain. In fact, as various others have argued (see, for example, Churchland 1986; Bechtel and Mundale 1999; Smolensky and Legendre 2006), the implementational details are often among the most explanatorily salient aspects of an account. Even if abandoning representational posits led to a greater sympathy for "ruthlessly reductionistic" (Bickle 2003) approaches, that would be a good result insofar as it broadens our conceptual and theoretical resources for understanding cognition.

Finally, how should we think about the cognitive revolution, which is so often framed as involving the introduction of representations as explanatory posits? While it is far from clear that this characterization is fully correct—there were many neo-behaviorists like Tolman (1948) and Hull (1930) who posited intervening states between stimuli and response that resemble modern notions of representation (Ramsey 2007)—there is no doubt that the new cognitivist picture of the mind involved the acceptability of positing representational states and processes. But making representations acceptable is not the same thing as requiring them. We can make perfectly good sense of the role representations played in the revolutionary developments in psychology in the mid-twentieth century without at the same time claiming representations are an essential element of all cognitive processes or must be posited by any serious model of cognition. Appealing to non-representational processes to explain some area of cognition shouldn't have the result that the theorist is suddenly labeled as a non-cognitivist or behaviorist.

What I am suggesting here is that we come to think about the relation between cognition and representation in the way we have come to think about the relationship between cognition and consciousness. At one time, mentality was simply equated with conscious thought and conscious experience. For earlier thinkers like Descartes, it simply didn't made sense to talk about unconscious or non-conscious mentality and conscious mentation defined the subject matter of the discipline of psychology. This equation of mind with consciousness was rooted in a traditional and commonsense conception of the mind. Yet today we have come to recognize that the consciousness demarcation criterion for mentality was misguided. We now acknowledge that there are many sorts of cognitive processes and psychological states that are not conscious and thus, that consciousness is not a defining characteristic of mentality.[1] We recognize that while consciousness is an important dimension of many cognitive states and processes, it should not be a demarcation criterion. By removing an artificial restriction on our conception of mentality—one that was not based on sound scientific reasons—we have allowed cognitive science to develop in ways that are driven by legitimate theoretical factors. For example, we now grant that many non-conscious neurological operations are legitimate topics of *psychological* research because they are functionally essential

[1] For a dissenting voice, see Searle (1992).

stages of more conventional psychological processes like memory or perception. What I am suggesting is that we should now adopt the same attitude with regard to non-representational operations. We should broaden our basic conception of cognition to include mechanisms and processes that have no representational states or structures whatsoever.

3.2 Argument 2: RDT undermines the representational theory of mind

The representational theory of mind is widely regarded as a contingent, empirical hypothesis about the underlying mechanisms of cognition. In one popular form, it is an attempt to explain the alleged intentionality of cognition by claiming that states like propositional attitudes are types of relations between cognitive agents and internal representations. But it also comes in the form of various computational theories of mind that posit various representational processes. As with our understanding of information processing, exactly what counts as "computational operations"—whether they are supposed to be the syntactic manipulation of discrete symbols or the spreading activation of nodes in a network—is not terribly restrictive. But whatever the nature of the processes, it is hypothesized that they involve representations and thus representationalism is presented as a *hypothesis about* the underlying architecture of cognition.

Yet none of this make any sense if we embrace RDT. After all, how can representationalism be *both* a falsifiable explanatory theory about the building blocks of cognition, and, at the same time, an essential, defining criterion of cognition itself? How can the positing of inner representations be an explanatorily novel and illuminating feature of theories that are supposed to account for various cognitive processes, and simultaneously be a necessary, qualifying condition for *theorizing about* cognitive processes? You can't treat representational posits as *both* interesting explanatory constructs *and* as a necessary condition for a legitimate account of the phenomena you are trying to explain.

How did we come to treat an allegedly empirical hypothesis about mental phenomena as a way of defining mental phenomena? I suspect part of this is due to the dual role representation plays in our psychological theorizing. As we've seen, it functions as an explanatory posit in various information-processing accounts of cognition. But it also functions as part of the *explanandum*—as a defining element of the very phenomenon we want to understand. After all, mental representations are at the center of conventional descriptions of certain cognitive capacities that we are investigating. Activities like believing, remembering, reasoning, imaging, and so on, are all cognitive activities that, at least prima facie, are inherently representational in nature and that researchers are attempting to explain. Thus, there may be a conflation of the explanandum with the explanans. That is, I suspect people notice that certain core cognitive phenomena we want explained, like memory and thinking, are representational phenomena, and thereby take these to more or less define cognitive phenomena. This in turn gets twisted into a claim not just about cognition-qua-explanatory-target, but also about what cognitive modeling and theorizing needs to be. But if this is what's going on, it is fundamentally confused. First, insofar as our basic conception of *some* cognitive processes or states entails that those processes or states fundamentally representational in nature, it doesn't follow that they are all that way. While things like memory

may be fundamentally representational, other cognitive capacities, like certain motor skills or types of learning, may not be. Second, insofar as representations form part of the explanandum of some cognitive processes, it doesn't follow that they should also serve as part of the explanans for all, some or even *any* cognitive processes. In fact, explaining representational phenomena like mental imagery by positing *other* internal representations always runs the risk of circularity or regress. And even if the best accounts of certain cognitive phenomena invoke inner representations, this in no way suggests that they all should. There is nothing problematic about claiming that representational cognition is only a sub-species of a more generic category that also includes non-representational cognition. Which cognitive processes are best explained by invoking representations and which ones aren't is an empirical question. Consequently, a representational theory of, say, our ability to perform some behavior should be a substantial, falsifiable hypothesis that could, at least in principle, compete with non-representational theories of the same phenomena. It shouldn't be a winner by default simply because the latter doesn't qualify as a cognitive account.

Again, our treatment of consciousness in cognitive science provides a helpful model. On the one hand, conscious access is sometimes invoked as part of an account of how we perform certain cognitive tasks, such as certain forms of perception or learning. As such, consciousness forms part of the explanans. But on the other hand, consciousness is itself a core aspect of many cognitive processes that we are trying hard to understand. So, like representation, consciousness (or introspective access) also plays a dual role as both part of certain explanatory proposals and also as an explanatory target. Yet the fact that consciousness is a critical element of some cognitive capacities we are trying to understand in no way entails that there cannot be non-conscious cognitive processes, or that conscious states must always be part of an explanation for some area of cognitive science. The fact that consciousness is clearly an important dimension of cognition does not lead us to think that it is a defining feature of cognition. This is how we should also think about representation.

3.3 Argument 3: it encourages wildly deflationary accounts of representation

If a discipline treats the positing of an explanatory construct as a necessary condition for proper theorizing within that discipline, then we can expect to find a significant cheapening of that explanatory construct. Theorists will stretch the notion so that it can be made to apply to the specific structures and states that they hypothesize so that their accounts can have legitimacy within the discipline. As I have argued at length elsewhere (Ramsey 2007), this is exactly what has happened with the concept of representation in the discipline of cognitive science. Classical computational accounts of cognition invoke symbolic structures that function in a recognizably representational manner by, for example, modeling various aspects of an environment or target domain. But along with classicalism, there are other sorts of cognitive architectures and hypotheses that posit their own types of internal structures. Among these is connectionism, which, as noted above, is also regarded as a representational framework. And yet when we look at the functionality of proposed structures described as representations within many connectionist models, we don't see anything recognizably

representational in nature. The internal nodes and connection weights are certainly critical for the processing, but there is no legitimate reason to describe their mediating role of transmitting waves of activation as one of representing anything. So why are they treated as representations? I suspect it is at least partially because their proponents want these models to qualify as cognitive models, and in the current climate, that requires internal representations. Note, for example, how Aydede suggests that without representations, connectionism would no longer qualify as a cognitive theory:

> …it is not clear at all how connectionism can genuinely give support to intentional eliminativism as far as the units (or collections of units) in connectionist networks are treated as representing. If they are not treated as such, it is hard to see how they could be models of *cognitive* phenomena, and thus hard to see how they can present any eliminativist challenge. (Aydede 2010, emphasis in original).

Here Aydede presents a clear expression of RDT—unless the nodes in connectionist models are serving as representations, the models can't really be viewed as models of cognitive phenomena. In such a climate, it is not surprising that the notion of representation has been deflated to the point where just about anything that is causally activated by input to a system and plays a mediating role in the processing gets described as a representation. The representational theory of mind is slowly becoming the "causally relevant to the processing theory of mind"—an utterly vacuous outlook. Yet this development is exactly what we should expect from a methodological stance that sanctions only representational theories as legitimately cognitive.

I should reiterate that my primary goal here is not to beat up on representations in general. What I am instead pointing out is that both anti-representational skeptics and staunch representational supporters should insist upon conceptions of representation that are legitimate, where the posit in question is doing something recognizably representational in nature. Because RDT encourages very weak and empty conceptions of representation, it should be rejected by all sides. Indeed, if you are strong proponent of the explanatory role of representations in cognitive science, then you should strongly oppose RDT. Not only does it undermine the empirical nature of representational theories, as we saw above, but it also severely damages the meaningfulness of the notion of representation.

Once again, the way we think about consciousness in cognitive science is instructive. There have been various reductive, deflationary accounts of consciousness that define it as something like the integration of information, the monitoring of internal systems or the control of behavior. Yet these are often derided as ignoring the so-called "hard" problem of explaining actual phenomenal experience—in other words, what we tend to think of as real consciousness (Chalmers 1996). This is not to say that explaining information integration or bodily control are unimportant. But we recognize the need to retain a very robust conception of consciousness, especially where it is supposed to do real explanatory work as, say, in the experience of subjective moods. We should also retain a robust conception of representation if it is to have any real explanatory value. The best way to do so is to stop treating representation as a necessary condition of cognition.

To summarize, the representation demarcation theis (RDT) is a common tendency to demarcate the cognitive in representational terms. I've offered three arguments against doing this: (1) it unnecessarily restricts our theorizing about cognition, (2) it undermines the empirical nature of the representational theory of mind, and (3) it encourages substantial weakening of the notion of representation. But if we are not going to demarcate the cognitive from the non-cognitive by appealing to representations, then how are we going to do so? In the next section I will explore an answer to this question that strikes me as the most sensible.

4 The non-representational conception of cognition

If we don't appeal to representations in defining what counts as cognitive, then what should we appeal to? I recommend that before trying to answer this question, we first pause to consider whether or not it actually needs answering. After all, it is not at all obvious why this is something we should care deeply about. Just why is the label "cognitive" so important and what exactly is at stake in our efforts to demarcate cognition? Suppose we wind up with an explanatorily satisfying neurological account of some dimension of the mind we want to understand. Does it really matter whether or not we can call this account a genuinely cognitive one, as long as we wind up understanding the things we want to understand? Is there anything, beyond turf wars and funding issues that demands articulating a clear set of criteria by which we should separate the cognitive from the non-cognitive?

I'm inclined to think not. I'm inclined to think most of the time there really isn't a lot at stake, scientifically, in our efforts to delineate the conceptual boundaries of cognition. If we were to go about trying to understand different neurological and psychological phenomena and processes without worrying about which ones really count as cognitive and which ones don't, not a lot would be lost. Still, at the same time, I don't expect people to stop trying to categorize the discipline of cognitive science, or to stop searching for the defining features of cognition. Hence, it is worth thinking about a more realistic and pragmatic approach to the demarcation problem. I suggest that instead of seeking strict classification criteria, we strive for something less stringent and more amenable to revision and modification.

For starters, I believe we should stop trying to define cognitive science by appealing to explanatory strategies, methodologies, or theoretical commitments. We should stop trying to define cognitive processes by suggesting they must involve specific sorts of states or particular kinds of operations. We should instead do what we do in practically every other intellectual discipline; namely, characterize a field of study in terms of what it is that is being studied. That is, we should adopt the standard way sciences are defined—in terms of their relevant explananda, in terms of what it is we want explained. Geological phenomena and theories are different from chemical phenomena and theories because their subject matters are quite different, even though they can sometimes overlap. Roughly, geology deals with the formation of mountains and rocks and minerals and so on. We don't define a geological explanation by stipulating the explanation must entail certain theoretical commitments or that the theories must posit particular types of states and structures.

The same point should apply to the sciences of the mind. The problem with the representational demarcation of cognitive science isn't simply that it demands a theoretical commitment that is problematic. It is that it demands a theoretical commitment at all. We should instead think of cognition as a crudely defined cluster of capacities and mental phenomena that we are trying to understand. A theory is a cognitive theory if it helps us to understand a capacity or process or phenomenon that we are pre-disposed to regard as psychological in nature. What matters is not anything about the theory itself—not its explanatory strategy or posits or even methodological constraints. Cognitive science should not to be defined by the *type* of explanation being offered. Cognitive science should defined by the *target* of the explanation. It should be demarcated by the kinds of questions we are trying to answer, not by the sort of answer that is on offer.

Of course, this immediately brings us back to the question of what should count as a cognitive process or cognitive phenomenon. If we are defining cognitive theorizing and cognitive science as theorizing about cognitive operations and states, then what exactly are cognitive operations and states?

Here I believe that we can and should lean heavily upon our intuitive, pre-theoretical image of the mind. Our ordinary conception of mentality is in all likelihood a cluster concept with fuzzy boundaries, with some prototypical processes in the center and more obscure or atypical processes and states on the periphery. We should use this common-sense and folksy conception of psychological capacities as a starting point, and then refine our thinking as the need arises. If, say, there are brain processes that are critical for storing memories, then because memory is a crucial aspect of our folk conception of the mind, we should treat those brain processes as cognitive in nature, and theories about those brain processes as cognitive, irrespective of the theoretical entities they invoke or the types of operations they describe. By contrast, if there are other brain processes that govern, say, our normal heart rate, and if the regulation of our heart rate is not something we pre-theoretically think of as a type of psychological process, then we should not regard those brain processes as cognitive in nature. Thus, I'm suggesting we let our pre-theoretical outlook serve as a rough guide for capturing the sort of capacities and states we should call cognitive, with a willingness to expand or shrink that outlook as the science demands.

An obvious concern the reader might have about my proposal is the following. Earlier I argued that a problem with RDT is that it restricts our theorizing in a way that is problematic because it is grounded in our pre-scientific, folk conception of the mind, and that folk conception may well be wrong. I argued that, as with folk physics or folk biology, we should not let our folk psychology overly constrain the way we define cognition. But now it seems I am doing precisely that by recommending we let our pre-theoretical, folksy conception of mentality guide our understanding of the subject matter of cognitive science. By appealing to our pre-theoretic conception of mentality, I seem to be doing exactly what I claimed earlier is a mistake in thinking about cognition.

In reply to this worry, let me emphasize that what I am recommending is substantially different from what I complained about earlier in at least two important ways. First, I am not suggesting that we lean on our commonsense understanding of the mind to establish the kind of states, structures or processes a theory must invoke to explain

cognition. I am not claiming we should make our scientific psychological explanations look just like our folk psychological explanations of behavior. Instead, I am suggesting we use our folk conception as a rough guide for thinking of the *subject matter* of cognitive science. While folk psychology should not restrict our thinking about what counts as a good theory, it can and probably must guide our thinking about what a good theory in psychology is a theory *about*. In fact, this is what happens in practically every discipline. Geology is at least initially defined as being about, commonsensically, rocks and mountains and minerals and that sort of stuff. Biology is more or less about living types of stuff and phenomena, pre-theoretically conceived. In nearly all intellectual disciplines, we use a rough commonsensical conception of the world to more or less figure out the intuitive joints of nature. We then use this rough picture to delineate different fields of study. As a starting point, I see nothing wrong with this.

Second, I am not suggesting that we use our folk conception of the mind to determine hard and fast criteria that trump other considerations when thinking about cognition. I am instead suggesting that we use it as only a rough guide for thinking about psychological phenomena—one that is always revisable in light of new discoveries and scientific developments. This isn't just because our folk conception of the mind is itself ill-defined with its own vagueness and imprecision. It is because it is also fallible and untrustworthy and should not be taken too seriously. If down the road we discover legitimate scientific reasons for treating the regulation of our heart rate as a cognitive process, then by all means we should include it as part of the subject matter of cognitive science, irrespective of how counter-intuitive this might now seem. The concern I raised earlier was a concern about rejecting theoretical proposals and research avenues if they weren't sufficiently compliant with our pre-theoretical beliefs. The way to avoid this worry is to not give those pre-theoretical beliefs veto power.

To some degree, what I am suggesting here is what a lot of people already do when they are forced to designate the discipline of cognitive science. For example, as anyone who has taught a course on cognitive science knows, or has edited an introductory volume on the subject, there is an intuitive set of psychological capacities, processes, states, and operations that we recognize as what cognitive science is about. Learning, perception, memory, reasoning, action, language, consciousness and so on….these are the sort of things we associate with the mind and thus also associate with cognition. People working to improve our understanding of these capacities are working in the field of cognitive science, regardless of the explanatory strategies they adopt or their specific theoretical commitments.

It is worth considering how the perspective I am endorsing ties in with some other recent trends in cognitive science. As I noted earlier, there are several recent writers who promote skepticism about representation because they reject, in various ways, their explanatory value. For example, those who endorse the perspective that has come to be known as "Radical Embodied Cognition Science" (REC), such as Chemero (2011) and Hutto and Myin (2012) offer various arguments against the theoretical usefulness of positing representations to explain a broad range of cognitive capacities. Similarly, proponents of the Dynamic Systems approach to understanding cognition, like Thelan and Smith (1994), van Gelder (1995) and Beer (1995) also propose a radical shift away from the information-processing perspective, in favor of mathematical models that plot the system's dynamics over time. Dynamicists tend to reject not

only computational models of cognition, but also the representational processes that underlie those computations.

Now as I have emphasized, my goal here is not to endorse any sort of across-the-board representational skepticism. However, I do want to make room for such radical new approaches in the discipline of cognitive science. As I see it, the only way that is possible is if we abandon RDT as I have recommended here. Thelan and Smith, for example, count as working in the field of cognitive science, not because they employ a representation-laden, information-processing account of how humans learn to walk (they don't), but simply because they offer a scientifically intriguing theoretical framework for explaining that cognitive achievement. Their framework may not ultimately pan out, but it certainly belongs alongside other novel theoretical proposals in the cross-disciplinary field of cognitive science.

Returning to an earlier point, what should we say about behaviorist and neo-behaviorist approaches to understanding psychological capacities? A worry about my proposal is that if what counts is the attempt to explain cognitive phenomena, and the theoretical commitments and explanatory strategies are irrelevant, then why shouldn't we say that behaviorists like Skinner and Watson were really cognitive scientists? After all, they were trying to provide a scientific explanation of such cognitive phenomena as learning and problem-solving. Proponents of RDT will claim this is exactly why the theoretical commitments *do* matter, contra the "big tent" proposal suggested here. They would balk at a conception of cognitive science that includes misguided behaviorists.

In response, it should first be noted that behaviorists, to some degree, committed the same sort of error I am attributing proponents of RDT, namely, defining a discipline by appealing to theoretical commitments and explanatory assumptions. Whereas RDT makes the invoking of representations a condition for legitimacy, behaviorists did just the opposite—insisting that psychological science must *not* posit inner representational states and processes. On my view, it is just as wrong-headed to preclude the invoking of certain theoretical posits as it is to require it.

Second, I concede that according to what I claim is the proper outlook, it is hard to see why early and modern behaviorists shouldn't also count as people working in the field of cognitive science. They are indeed trained scientists trying to understand important cognitive phenomena, albeit in a way that many regard as deeply misguided. But misguidedness should not be a disqualifier here or in any other scientific discipline. So I am prepared to bite the bullet and recommend treating them as contributors to the field of cognitive science, as people studying important areas of cognition. Of course, that doesn't mean we can't, in many ways, treat their methodologies and assumptions as radically different from more mainstream cognitive investigators. We can still characterize traditional psychological behaviorists as anti-representationalists or as 'anti-inner processes' as many now do, and thereby distinguish them from those who pursed the information-processing approach that came later. But I see nothing terrible about claiming the so-called "cognitive revolution" was a revolution in the *way* cognition was investigated and explained, not in the commencement of cognitive research. Moreover, it should be noted that the proponent of RDT has the opposite embarrassment—an extreme form of exclusivity. RDT proponents are forced to insist that renowned and prize winning researchers like Randall Beer and Linda Smith aren't

really cognitive scientists—despite having their home in a cognitive science program and running a Cognitive Development Lab at Indiana University—simply because they eschew the representational/information-processing approach in their exploration of cognition. That strikes me as far more bizarre than treating people like Skinner as primordial cognitive scientists.

In philosophy, another recent and related trend concerns what counts as the actual physical boundaries of cognition. Proponents of the "extended cognition" hypothesis, like Clark and Chalmers (1998), argue that cognitive systems should extend beyond the biological nervous system—that, say, a notebook used to store information for someone with a memory deficit should, in the right circumstances, be viewed as that person's actual memory. Thus, according to them, a cognitive system includes the biological brain *and* various non-biological devices that perform information processing and storage duties, serving not as aids to the cognitive system, but as parts of it.

Those opposed to the extended cognition proposal have argued against it in one of two ways. On the one hand, some have specified necessary conditions for parts of cognitive system, and then argued that non-biological elements lack this condition. For instance, Adams and Aizawa (2001) argue that real cognitive components possess representational states with original or non-derived intentionality (that is, the meaning of the representations does not depend upon any sort of outside interpretation). Since things like notebooks lack symbols with original intentionality, they do not qualify as cognitive. On the other hand, some have argued against extended cognition by appealing to the ways in which the parts of cognitive systems must be properly integrated and conjoined. Rupert (2009), for example, defines a cognitive state as something that is part of a mechanism that causally contributes to the production of different cognitive phenomena in a systematic way. He then argues that things outside of the nervous system, like notebooks, are not integrated in this way, and thus should be viewed as aids to (as opposed to parts of) the cognitive system.

How does all this mesh with my proposal? On my view, there is nothing wrong, in principle, with modifying our conception of a cognitive system in the way Clark and Chalmers suggest, as long as it is properly motivated by legitimate considerations that will help our understanding of cognition. However, at the present I do not see how such a revision would help; that is, I don't see how the extended outlook on the mind is superior to our more intuitive conception of a cognitive system as (roughly) the brain and nervous system of cognitive agents. As far as I can see, there really aren't any compelling scientific or explanatory reasons for shifting our understanding of cognitive systems in this way. As for the suggestion that cognitive processes must involve internal states with a certain feature like original intentionality, as Adams and Aizawa suggest, this is precisely what I believe we should *not* do. In fact, Adams and Aizawa's original intentionality condition for a cognitive system is just a variant of RDT. Their proposal would restrict our understanding of cognition to the representational in exactly the way I've argued we shouldn't.

On the other hand, Rupert's system-based proposal for defining cognition and cognitive systems converges rather nicely with the proposal offered here. Intuitively understood cognitive capacities and processes are explained by the causally relevant mechanisms and operations that are integrated and interconnected in a manner that forms a cognitive system. Any theory about the workings of those mechanisms is thus a theory

about the nature of cognition. Rupert's account is properly unrestrictive and flexible in just the way that should allow for unfettered cognitive modeling and theorizing; it also provides a plausible method of distinguishing internal cognitive components from external cognitive accessories. This system-based criterion for defining the physical borders of cognitive systems is fully compatible with my proposed characterization cognitive processes and cognitive theorizing.

Finally, one further new development in cognitive science is a growing appreciation for the explanatory value of hybrid models or different forms of architectural pluralism. For example, one popular development in cognitive modeling is often referred to as the "dual-process" or "dual-system" picture of cognitive architecture (Evans 2010; Evans and Frankish 2009; Frankish 2004; Sloman 1996; Stanovich and West 2000). As the label implies, proponents of this approach treat the mind as composed of two types of architectures that function in fundamentally different ways. One system, S1, is claimed to perform cognitive operations that are non-conscious, fast, and automatic. It performs such tasks as pattern recognition and associative reasoning. A second system, S2, is assumed to operate in a way that is slower, more deliberate, and more rule-governed, supporting our conscious thought processes. On some versions, distinct cognitive capacities like procedural knowledge or logical inference are thought to be housed in either S1 or S2, whereas in other accounts, both systems are thought to contribute something to basic cognitive processes like reasoning or memory. Space does not allow a detailed examination here, but there is one aspect of this picture that directly bears directly upon our discussion.

Insofar as the S1 architecture may sometimes involve more "primitive" or "lower-order" forms of processing, there may be little need for sophisticated information storage or full-blown representational mechanisms or processes. Automatic processes are often characterized as those that are purely causal and occur without the retrieval of information-bearing states. Now according to RDT, such processes would not really qualify as cognitive in nature. Despite their relevance to various higher-level capacities (like recognition or skill learning) and despite their being housed in what is often characterized as a type of cognitive architecture, RDT would nonetheless force us to treat these operations as non-cognitive (or perhaps sub-cognitive) given that they lack representations. That seems silly. A far more reasonable attitude toward the dual-architecture approach would be to treat both S1 and S2 as types of cognitive architectures involving cognitive mechanisms and cognitive processes of different sorts. Whereas many of the processes in S1 may be rich with representational states and processes, they need not be. S1 is often characterized as being a cognitive architecture that is, more or less, non-conscious, non-rule-driven, non-deliberate, non-sequential, and so on. I propose that we also be willing to include one more on the list of "nons": non-representational.

5 Conclusion

I have argued here against the common practice of demarcating cognition and cognitive theorizing in representational terms. For some anti-representationalists, my arguments should provide indirect support for their unconventional outlook on cognitive science.

But even for those who embrace the explanatory role of representations in cognitive science, it is important to recognize that a healthy science of the mind should not restrict itself by always requiring certain explanatory posits like inner representations. Such a requirement places an artificial boundary on our theory-development, undermines the explanatory strength of representations and encourages bad science. Severing the conceptual tie between cognitive science and representation will improve our investigation and understanding of both.

Acknowledgements Versions of this paper were presented at the "What is Cognition Conference?", Center for Mind, Brain and Cognitive Evolution, Ruhr-Universität Bochum, Germany, June, 27–29, 2013; the "Reach of Radical Embodied Cognition Conference" University of Antwerp, Belgium, June 17–19, 2013; and the University of Milan, Italy, June 21, 2013. I am very grateful for helpful feedback from all of these audiences. I am also grateful to Ken Aizawa, Cameron Buckner, Ellen Fridland and two anonymous referees for extremely helpful suggestions and recommendations.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*(1), 43–64.
Aydede, M. (2010). The language of thought hypothesis. In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy (fall 2010 edition). Retrieved from http://plato.stanford.edu/archives/fall2010/entries/language-thought/.
Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, *66*, 175–207.
Beer, R. D. (1995). A dynamic systems perspective on agent–environment interaction. *Artificial Intelligence*, *72*, 173–215.
Bermudez, J. (2010). *Cognitive science: An introduction to the science of the mind*. Cambridge: Cambridge University Press.
Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. New York: Kluwer/Springer Publishing.
Brentano, F. (1874). *Psychology from an empirical standpoint*. London: Routledge and Kegan Paul.
Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.
Chemero, A. (2011). *Radical, embodied cognitive science*. Cambridge, MA: MIT Press.
Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.
Evans, J. S. B. T., & Frankish, K. (2009). *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.
Evans, J. S. B. T. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.
Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.
Frankish, K. (2004). *Mind and supermind*. Cambridge: Cambridge University Press.
Hull, C. L. (1930). Knowledge and purpose as habit mechanisms. *Psychological Review*, *37*, 511–525.
Hutto, D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: MIT Press.
Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.
Newell, A., & Simon, H. (1976). Computer science as empirical inquiry. *Communications of the ACM*, *19*(3), 113–126.
Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*, 135–183.
Ramsey, W. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.
Rowlands, M. (2010). *The new science of the mind*. Cambridge, MA: MIT Press.
Rupert, R. (2009). *Cognitive systems and the extended mind*. New York: Oxford University Press.
Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22.

Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 1), Cognitive architecture Cambridge, MA: MIT Press.

Stanovich, K., & West, R. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–665.

Thagard, P. (2014). Cognitive science. In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy (fall 2014 edition). http://plato.stanford.edu/archives/fall2014/entries/cognitive-science/.

Thelan, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.

van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *91*, 345–381.