

Disentangling Representation from Conceptualisation

Nancy Adrienne Salay (salay@queensu.ca)

Queen's University, Department of Philosophy, Watson Hall 309
Kingston, ON K7L 3N6 CANADA

Abstract

Drawing on recent work in the area of episodic memory, I suggest a novel way of dissolving the representation/anti-representation debate; if we treat representation and conceptualisation as two separate capacities, the latter being parasitic on the former, we unify the insights of both camps, but succumb to none of their failings. I provide a sketch of how we might understand de-conceptualised representation and I show that, on this new approach, many of the old problems, e.g. grounding, disappear.

Keywords: representation; anti-representation; intentionality; episodic memory; cognitivism; cognition; dynamic systems

Introduction

One of the central debates in cognitive science is the dispute over the role of representation in cognition: on one view, cognition is taken to be a kind of conceptual process; on the other, it's seen as a particular sort of physical process. But this characterisation of the two perspectives leads to a deeply unsatisfying theoretical divide: either our interactions with our world are mediated by our representations or we do a lot less modelling than we think. Accounts situated in the representational camp are plagued by the problem of intentionality, while those leaning towards the anti-representational side seem incapable of saying anything theoretically interesting about the higher-level conceptual type of cognition that purportedly develops out of the underlying non-representational physical processes. This unhelpful polarisation is due, I think, to a presupposition that drives the debate; the idea that representations are conceptual¹, that is, that they stand in for classes of things, not for particulars. I want to suggest that if we unpack this, and treat representation and conceptualisation as two separate capacities, the latter being parasitic on the former, we get a more nuanced and robust theory of cognition that unifies the insights of both camps.

In what follows, I'll take Hubert Dreyfus' arguments against cognitivism (computational/representational approaches) as paradigmatic of the anti-representational camp, since his are the most explicit and best worked out in the literature. I'll describe the competing views, uncover the underlying intuitions that motivate the respective

positions, and then suggest a new paradigm for understanding mental re-presentations as symbols of particulars.

Why Cognitivism is in Trouble

The central point of contention between Dreyfus and the cognitivist is whether or not cognition is ultimately a matter of *symbol* manipulation. Cognitivism begins with the idea that a cognitive system responds intelligently to its environment *by way of* its internal symbols but it is precisely this commitment, Dreyfus thinks, that constitutes cognitivism's fatal flaw. As soon as one thinks that internal stuff can stand for external stuff, the problem of how the symbols are connected to their objects arises. What makes the symbols *symbols*? How do the inner representations get to be *about* the things they represent? What makes the inner models, *models* of the external world and, further, models of this specific part of the external world? If, as Quine so eloquently argued (1951), concepts are so deeply entangled with one another that having one entails that one must also have a host of others, this web of concepts – even if internally coherent – could never model the stuff out there; such a system could never decide, for example, that a particular real-world situation is a birthday party situation because the rule for recognising a birthday party situation would cite other concepts that would in turn need rules of application and those rules would also cite concepts that would need rules of application and so on *ad infinitum*.

Now of course one way to hold on to the representational story, but avoid the infinite regress, is to reject concept holism and suppose that there is some level of symbols for which the fixing process is not mediated by other symbols, that is direct in some non-representational way. This is the project of giving a naturalistic account of intentionality. If all holistically inter-linked symbols are 'grounded' in a level of symbols that have their contents fixed in some non-intentional way, the idea goes, the regress will end there. Recognition that representational accounts that don't have some grounding layer – we should think of these as disembodied or ungrounded cognitivist approaches – face serious problems of exactly the sort that Dreyfus raises has made the project of "solving the grounding problem" an increasingly topical one.² But, for Dreyfus, grounding cannot be a way out of the vicious circle since grounding requires a commitment to some form of concept atomism – the view that individual symbols can represent

¹ Indeed, Fodor (1990) goes to some length to argue that being a generalisation is a *requirement* for being a representation; symbols that stand in for particulars only are mere "labels" on his view. I think this is deeply mistaken and rests on the confusion that is the focus of this paper.

² See Taddeo & Floridi (2005) for a good review of the current state of work in this area.

independently of all other symbols – and he thinks that there are independent reasons for thinking that concept atomism must be false.

Dreyfus thinks that concept atomism is wrong in two ways. Not only does it not speak to the Quinean intuitions about concept holism – that concepts come in groups, not as individuals – but it assumes that our most basic relation with the world is a conceptual one. Dreyfus thinks that both the atomists and (some) holists are wrong in thinking that, at base, we come to know our world by theorising about it. He follows Heidegger in insisting that fundamentally our relation to the world is not a theoretical one, but a practical one. On this Heideggerian view, the idea that *things in the world* are meaningful in isolation from our *practices* is incoherent. Things are what they are and play the role they play partly because of their natural features – being heavy, being sharp, and so on – and partly because of the background practises of the culture within which those roles develop.

Although practical understanding – everyday coping with things and people – involves explicit beliefs and hypotheses, these can only be meaningful in specific contexts and against a background of shared practices. And just as we can learn to swim without consciously or unconsciously acquiring a theory of swimming, we acquire these social background practices by being brought up in them, not by forming beliefs and learning rules. (Dreyfus, 1980, p.7)

Practical understanding underwrites and makes theoretical understanding possible: meaning is always situated, that is, it arises out of holistic, dynamic, inter-relations between agents and their environment. Meaning *arises*, as a whole, out of activity, and never individually, as a result of assignments:

To say a hammer has the function of being for hammering leaves out the defining relation of hammers to nails and other equipment, to the point of building things, and to our skills – all of which Heidegger called readiness-to-hand – and so attributing functions to brute facts couldn't capture the meaningful organization of the everyday world. (Dreyfus, 2008, p.1138)

Thus, it's the complex nexus of background practices *and* (sometimes) conceptual relations that holds together hammers, nails, wood, etc., that connects *hammer* with hammers. Any view that *requires* a base-level of concept-detectors, as it appears cognitivism does if it is to avoid the regress, has been completely derailed. The fact that it was the initial assumption of representationalism that was responsible for this flight towards conceptual atomism, Dreyfus urges, should be a red flag that something is deeply wrong with that assumption.

Now some might think that Dreyfus is creating much ado about nothing, that if there is any debate here it's merely a terminological one since the internal states, or at least some subset of them, that underlie our practical understanding just

are the low-level representations that ultimately ground a higher-level theoretical conceptual structure. The rising influence of neuroscience in the cognitive sciences coupled with the widespread acceptance of some kind of information theoretic account of representation has made this idea that conceptual cognition might be grounded in a simple capacity for object detection a natural one.

But this is no mere terminological conflict; it is a deep and confounding burden of proof debate: there seems to be no non-question-begging way of specifying what constitutes a representational system. If we are loose with our use of the term “representation” and we suppose that nomic covariance relations are sufficient to establish representation relations, then we are in danger of begging the question in one way, of assuming that *using a representation* and *acting in a way that can be interpreted as using a representation* are two sides of the same coin. This “loose” understanding is mainstream in neuroscience. When a neuron or a cluster of neurons is found to be ‘sensitive’ to a particular class of objects, and the underlying explanation is taken to be that a nomic causal relation between the objects of some class and the activation of a neuron or cluster of neurons has been found, that neuron or cluster is said to represent that class. There is a dispute, to be sure, over whether or not localist, single-cell, representations are possible or whether neural representations are distributed over clusters of neurons; but there is very little discussion (except among the anti-representationalists of course) about whether or not nomic covariance is sufficient to warrant representational attribution³.

But surely, one might think, there is a difference between what I do when I consult a map in order to find the shortest route across the city and what I do when I follow a series of instructions for crossing the city. In the first case I am using the map in virtue of its content, but in the second case, while the entire sequence of steps taken together could be viewed as a model of the shortest distance across the city, I do not follow the instructions in virtue of their semantic content, I follow them in virtue of their syntactic properties – turn left at Bank Street, proceed for two blocks, and so on. If I didn't understand the semantic features of the map, e.g. if I didn't know that the black lines represented streets and the red lines stood for highways, I wouldn't be able to use the map; on the other hand, whether or not I understood that, taken as a whole, the sequence of steps represented the shortest path across the city, I could follow the instructions for taking that path. In the first case I am using the representation, but in the second I am merely acting in a way that could be interpreted as using a representation. Neuronal chain reactions, looked at from an investigator's vantage point, can certainly be interpreted as

³ See, for example, the recent debate – rekindled by Jeffrey Bowers (2009) – concerning localist vs. distributed neural representations. Nowhere in this discussion is the question of whether or not we should be calling these regularities representational at all raised.

representational, in the same way that the sequence of steps can be seen as a model of the shortest route across the city, but unless one has some kind of story to tell about how one part of the system, or perhaps the system taken as a whole, makes use of the content of those neural states, we have no reason to think that these neuronal impulses actually play a representational role. Of course, by stating the distinction in this way, I am also begging the question, in the other way, since to get to my conclusion one has to first assume that *using a representation* and *acting in a way that can be interpreted as using a representation* are different. This is why the debate about representation seems so intractable; the competing intuitions that undergird the various positions are so polarised.

William Ramsey does a good job of making the gap between these alternative perspectives explicit when he argues that any account of representation must meet what he calls the *functional specification challenge*: “a minimal requirement for a successful functional specification of any notion of representation is that the content—or, if you like, the fact that the representation has semantic content—be an explanatorily relevant fact about that state.” (2003, p. 129) In other words, one needn’t go so far as to show that a system is actually using a representation in order to make the case that the system is a representational one (as in my map example) – it’s unclear how one could ever give a naturalistic account of intentionality if this were the requirement – but one does need to provide a justification for treating a system as though it were using a representation; that is, the fact that the internal indicator states have some content must play *some* kind of role in one’s account.

Fred Dretske’s information theoretic account of representation (1988), perhaps the most robust and ambitious indicator theory of representation that has been offered thus far, looks like the best candidate for meeting this challenge. According to that account, what makes one internal state X a representation of some class of things or actions Y is the following:

1. The presence/absence of X covaries with the presence/absence of members of Y;
2. The co-variance is under-written by a nomic causal relation, that is, the presence/absence of members of Y cause or are a necessary part of the cause of the presence/absence of X; and,
3. The functional role of X, within the system within which it arises, is to carry information about the presence/absence of members of Y.

It’s condition three that makes this account a candidate for meeting the functional specification challenge, since it’s this requirement that makes the content of the purportedly representational state relevant to a complete description of its functional role in the system. Or so it seems. Ramsey argues that it doesn’t. To support his contention, Ramsey develops a distinction between *carrying information about* –

“possess[ing] states that could inform about other states of affairs” (2003, p.135) and *being an informer*— “be[ing] plugged into the right sort of system in the right sort of way, such that the relevant entailment relations are put to a very specific sort of use.” (2003, p. 135) It’s the latter that is required to meet the functional specification challenge, since only in such cases is the information actually playing some kind of role in the overall account. But in none of Dretske’s examples, Ramsey argues, is the *informer* condition met.

I won’t rehearse here Ramsey’s support for this claim since ultimately it’s not important that we be convinced of Ramsey’s conclusion; indeed, one of the morals of this paper is that so long as some of our key presuppositions about the nature of representation remain, we will never be able to solve this burden of proof debate. Ramsey’s insights are important, however, because they uncover the fact that information theoretic accounts of representation are convincing only if we assume a particular (impoverished, on one view) understanding of representation; consequently, we shouldn’t be optimistic that an information theoretic grounding account could ever settle the score.

But as I’ve already noted, for anti-representationalists like Dreyfus, the entire grounding agenda, information theoretic or not, is misguided ultimately because it cannot accommodate our dynamic nature as systems who are continually responding to and causing changes in our environment. Any view on which it makes sense to see coping skills as decomposing into finer and finer grained skills at dealing with object types, even those that are “action-oriented” or Gibsonian, is a representational view by Dreyfus’ lights and thus one that he rejects. On the representational view, our interactions with the world are mediated by categories and we see the world *as* divided up into hammers and tables and chairs. In coping, on the other hand, there is no “seeing as” at all. One copes with situation wholes, as unfolding happenings, rather than as composites of objects.

Dreyfus’ anti-representationalism is thus quite radical; he rejects *any and all* representational interpretations of the internal states that underpin our coping skills. There is no mere terminological argument here.

Re-Presentation: A New Model

Dreyfus’ deep and important insight into the way we think has led us to the following impasse: any disembodied or ungrounded AI founded on the principle that cognition is fundamentally a matter of concept manipulation will be caught in an infinite regress of concept consultation and, consequently, its concepts will fail to be about the things they purportedly represent. But the mainstream cognitivist response, to close the concept-world gap by grounding conceptual schemas, requires that we take concept atomism seriously. Dreyfus rejects this route and takes the fact that this looks to be the only way out of the infinite regress as a

clue that the initial representational assumption must be at fault. Others, who are more firmly rooted in the Cartesian tradition, are not as quick to reject the grounding possibility; they think that some kind of information theoretic atomism will eventually provide the answer. But, as Ramsey show us, low-level concept detectors alone can't provide us with a naturalistic account of representation, since they cannot play the required functional role of concepts or even proto-concepts unless they are already part of an intentional system, a system capable of *using* semantic content. Such grounding theories, instead of solving the problem, merely push it back a level. Dreyfus urges us towards an anti-representational AI, but I suspect the radical see-saw between the full-blooded representationalist and the strident anti-representationalist is a tug 'o war that no-one is likely to win, likely because, as Andy Clark and others (1994, 1997, 2002) have been suggesting all along, the truth lies somewhere in the middle. I suggest that we take Dreyfus' charge seriously, but explore other ways out of the impasse. His solution is to reject the foundational commitment to some form of representationalism outright, but I want to argue that even if we accept Dreyfus' arguments that cognition isn't fundamentally a conceptual process, we need not accept his more radical and less helpful conclusion that it is also not a representational process.

As evolved biological agents, there is no doubt that, as Dreyfus emphasises, we first and foremost cope with our environment in an entirely non-representational way – we avoid obstacles, recoil from harmful situations, and are drawn towards safe and pleasant ones in an unmediated way. But we're not *just* biological agents; we're cognitive ones as well. And as such we are able to respond not only to the intricacies of present situations, but to past and future ones as well. I am able to respond to the subtleties of the ebb and flow of traffic, while I'm driving my car, while at the same time, thinking about how my class went yesterday and considering ways in which I will do things differently or the same in next week's class. But while this ability does require that we have some capacity to re-present the past and imagine future situations, it need not require a conceptual ability, an ability to generalise beyond the specific cases to a class. I have in mind here the distinction, first proposed by Endel Tulving, between episodic and semantic memory. Episodic memories are re-presentations of past experiences (and imaginings of future ones), while semantic memories are conceptualisations of past experiences – they consist in the knowledge we distil from our experiences, that is, the generalisations we make on the basis of experience. Proust's *In Search of Lost Time* gives us wonderfully vivid descriptions of both. The narrator has an episodic memory of a particular moment in his childhood, when he'd tasted a bite of his aunt's lime-tea-soaked madeleine, and in this re-experiencing he tells us that "immediately the old grey house upon the street, where her [his aunt's] room was, rose up like a stage set to attach

itself to the little pavilion opening on to the garden which had been built out behind it for my parents (the isolated segment which until that moment had been all that I could see); and with the house the town, from morning to night and in all weathers, the Square where I used to be sent before lunch, the streets along which I used to run errands, the country roads we took when it was fine." (p.50) This memory is portrayed in stark contrast with his more conventional semantic memories of Combray, the village of his childhood summers, where his aunt lived: "Many years had elapsed during which nothing of Combray, save what was comprised in the theatre and the drama of my going to bed there, had any existence for me." (*ibid.*) The ability to have an episodic memory, then, is the ability to re-experience some situation not present, while to have a semantic memory is to have some capacity for generalisation. Tulving, both when he first suggested the distinction and today, sees episodic memory as parasitic on semantic memory:

Episodic memory is a recently evolved, late-developing, and early-deteriorating past-oriented memory system, more vulnerable than other memory systems to neuronal dysfunction, and probably unique to humans. It makes possible mental time travel through subjective time, from the present to the past, thus allowing one to re-experience, through autothetic awareness, one's own previous experiences. Its operations require, but go beyond, the semantic memory system. (Tulving, p. 6)

Martin Conway, however, has recently suggested an intriguing new way of understanding episodic memories as a tripartite structure only elements of which are entwined with semantic memories. His analysis provides us with a new way of understanding the relationship between mental representations and conceptualisation, one that can serve as the foundation for the theoretical bridge between our coping skills and our conceptual abilities.

On this new picture, episodic memories can be analysed into inter-related parts:

1. Episodic elements (EE's) – these are snippets of experiences, Proustian experience snapshots;
2. Semantic episodic memories (SEM's) – these are small sets of EE's grouped by a contextualising conceptual frame, for example, one's breakfast routine; and,
3. Conceptual episodic memories (CEM's) – these are groupings of SEM's by a higher-order conceptual frame, for example, a day at work.

Of these, the most basic are EE's:

Episodic elements are the most event-specific, most experience-near representations in long-term memory. They are often in the form of visual images (which may be the main representational format of episodic memories) and they represent moments of experience or summaries of moments of experience, particularly

and perhaps exclusively, moments of conscious experience. (Conway, p. 2308)

Importantly, on Conway's view, and contra the received wisdom, EE's are a-conceptual and conceptually *a priori*. This means that the capacity for semantic memory is not a requirement for having episodic elements; indeed, he suggests, the relation goes the other way – episodic elements are required for semantic memory.

An interesting question that then arises is: how can EE's be associated with conceptual frames in an infant's memory? One answer to this is that the ability to form EE's is hard-wired and functioning prior to birth.

Conceptual knowledge is abstracted from EE's. (p. 2312)

Conway here is suggesting that free-floating EE's, what I'm calling re-presentations, that aren't yet grouped and framed, might be the building blocks of our concepts. At some point in a human infant's development, EE's begin to be grouped. These groupings form 'proto-SEM's', the beginnings of conceptual frames. The mechanism for grouping is of course the big ticket question since, on this view, this just is the mechanism for conceptualisation. An initial suggestion is that both the temporal contiguity of EE's and closeness with respect to sensory attributes are likely factors in how EE's are grouped in long-term memory.

Obviously, more work needs to be done in this area before we can claim anything as bold as a theory of concept development, but in speaking to the insights of both the representational and anti-representational camps, EE's as representations play an important role in the cognitive story we have collectively been telling thus far. The anti-representationalist is motivated by the bottom-up observation that cognitive agents such as ourselves are, most fundamentally, dynamical physical systems and, as such, are best described in terms of the low-level mechanisms from which our higher level capacities emerge. That representations are the building blocks out of which our conceptual capacities emerge supports this picture since representations themselves are just responses to past experiences, neurally encoded and re-played and, in that sense, are no different from other coping responses. The representationalist, on the other hand, is motivated by the top-down observation that, unlike most physical systems, cognitive agents are able to respond to past events and possible future ones in addition to present situations. Here again, the insight is captured since a re-presentation is either a response to an experience that has already happened or is a playing out of a possible future one: "The temporal dimension in episodic memory extends then both backward and forward in time and we have recently termed this the *remembering-imaging window*." (p. 2307)

Recall that attempts to ground conceptual schemas directly in some kind of body-world relation (this is what the detector accounts try to do) can never resolve the clash of intuitions. Either the attribution of representation to

internal indicator states will be unwarranted, given a more robust notion of representation, or it will be justified only if we assume that the overall system is already intentional, which is the very thing we are trying to explain. Instead of responding to this observation by embracing the opposing intuition, as Dreyfus does, I'm suggesting a possible theoretical middle ground. We do rely on inner models of our environment in our interactions, but, as part of a low-primitive cognitive capacity, these models are wholly particular, not conceptual. There is no gap between a representation and the world that needs to be bridged (as in the case of conceptual representation); the re-presentation is just another situation to be (re)experienced. Representations do not threaten the underlying anti-Cartesian picture of ourselves as, ultimately, dynamic copers, since representations don't mediate our interactions with the world; although they do make mediation possible, since it is out of these snapshots that concepts develop. Finally, representations meet the functional specification challenge without begging the intentional question. Being a neuronal response and thus embedded within a network of connections, a re-presentation triggers other responses as well. As a first response, of course, not re-experienced, there is nothing representational about the underlying neuronal structure that encodes that response. But once that set of neural encodings is re-activated, it now has the role of carrying information about the original situation for the system to which it is being re-presented. Not just any set of neural encodings can count as representational then, no matter how detector-like they behave; only those in a system that is capable of re-presenting to itself count as representations, since only in such systems is the experience as a whole, that is to say, the content of the experience, playing a role.

Conclusions and Speculations

It sounds like we get to have our cake and eat it too. Is this too good to be true? Perhaps, but it certainly opens up some new avenues of investigation where now we seem to be stalled. The anti-representationalist seems incapable of offering a theory of cognition, distinct from a theory of, say, action, because he ignores our theoretical capacities; the representationalist, on the other hand, seems incapable of offering a theory of embodied cognition because she begins the cognitive account too high up – already at the level of concepts. If we let go of the idea that a representation must relate some particular to a class, we have a way of marrying the insights of either approach and moving forward with a new conception of intentionality: an intentional being is one that has the capacity to respond to its own response to some past experience. Intentionality, on this view, is a *prerequisite for* conceptual cognition; that is, giving an account of intentionality and giving an account of mental concepts are separate endeavours. And this is very happy news because until now we've had the proper order of investigation backwards; the thought was that once we

understood conceptual representation, the right story about intentionality would follow. But no naturalistic account of conceptual cognition is forthcoming, no surprise, since that story is parasitic on, not precedent to, an account of intentionality. This scaled back understanding of intentionality as the capacity for re-presentation, however, looks like a much more promising candidate for naturalisation. If we can manage that, we'll have, at last, a naturalistic grounding for a theory of conceptual representation.

References

Clark, A. & Toribio, J. (1994). Doing without representing? *Synthese*, 101, 401–431.

Clark, A. (1997). *Being there: putting brain, body, and world together again*. Cambridge: MIT Press.

Clark, A. (2002). Is seeing all it seems? Action, reason, and the grand illusion. *Journal of Consciousness Studies*, 9(5/6).

Conway, M.A. (2009). Episodic memories. *Neuropsychologia*, 47, 2305–2313.

Dretske, F. (1988). *Explaining Behavior*. Cambridge: MIT Press.

Dreyfus, H. (1980). Holism and Hermeneutics. *The Review of Metaphysics*, 34 (1), 3-23.

Dreyfus, H. (2002). Intelligence without representation – Merleau-Ponty's critique of mental representation. *Phenomenology and the Cognitive Sciences*, 1, 367–383.

Dreyfus, H. (2008). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Artificial Intelligence*, 171(18), 1137–1160.

Freeman, W. (2000). *How brains make up their minds*. New York: Columbia University Press.

Fodor, J. (1990). Information and representation. In P. Hanson (Ed.), *Information, language, and cognition*. Vancouver: University of British Columbia Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335-346.

Keijzer, F. (1998). Doing without representations which specify what to do. *Philosophical Psychology*, 11(3), 269–302.

Keijzer, F. (2002). Representation in dynamical and embodied cognition. *Cognitive Systems Research*, 3, 275–288.

Kirsh, D. (1990). When is information explicitly represented? In P. Hanson (Ed.), *Information, language, and cognition*. Vancouver: University of British Columbia Press.

Prinz, J. & Barsalou, L. (2000). Steering a course for embodied representation. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual changes in humans and machines* (51-77). Cambridge: MIT Press.

Proust, M. (1913-27). *Remembrance of Things Past. Volume 1: Swann's Way: Within a Budding Grove*. The definitive French Pleiade edition translated by C.K. Scott Moncrieff and Terence Kilmartin. New York: Vintage.

Quine, W.V.O. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, 60, 20-43.

Ramsey, W. (2003). Are receptors representations? *Journal of Experimental & Theoretical Artificial Intelligence*, 15:2, 125-141.

Taddeo, M. & Floridi, L. (2005). Solving the symbol grounding problem: a critical review of fifteen years of research. *Journal of Experimental & Theoretical Artificial Intelligence*, 17 (4), 419–445.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annu. Rev. Psychology*, 53, 1–25.

Van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, XCII(7), 345-381.

Varela, F., Thompson, E., and Rosch, E. (1991). *The embodied mind*. Cambridge: MIT Press.

Wheeler, M. (2008). Cognition in context: phenomenology, situated robotics and the frame problem. *International Journal of Philosophical Studies*, 16 (3), 323-349.