From Francisco Garzon and John Symons (eds.) The Routledge Companion to the Philosophy of Psychology, London: Routledge (2009): pp. 233-250.

Problems of representation I: nature and role

Dan Ryder

University of British Columbia

<3>Introduction</3>

There are some exceptions, which we shall see below, but virtually all theories in psychology and cognitive science make use of the notion of *representation*. Arguably, folk psychology also traffics in representations, or is at least strongly suggestive of their existence. There are many different types of things discussed in the psychological and philosophical literature that are candidates for representation-hood. First, there are the propositional attitudes – beliefs, judgments, desires, hopes etc. (see Chapters 9 and 17 of this volume). If the propositional attitudes are representations, they are person-level representations – the judgment that the sun is bright pertains to *John*, not a sub-personal part of John. By contrast, the representations of edges in V_1 of the cerebral cortex that neuroscientists talk about and David Marr's symbolic representations of "zero-crossings" in early vision (Marr 1982) are at the "sub-personal" level – they apply to parts or states of a person (e.g. neural parts or computational states of the visual system). Another important distinction is often made among perceptual, cognitive, and action-oriented representations (e.g. motor commands). Another contrast lies between "stored representations" (e.g. memories) and "active representations" (e.g. a current perceptual state). Related to this is the distinction between "dispositional representations" and "occurrent representations." Beliefs that are not currently being entertained are dispositional, e.g. your belief that the United States is in North America - no doubt you had this belief two minutes ago, but you were not consciously accessing it until you read this sentence. Occurrent representations, by contrast, are active, conscious thoughts or perceptions. Which leads us to another important distinction:

between conscious and non-conscious mental representations, once a bizarre-sounding distinction that has become familiar since Freud (see Chapter 4 of this volume).

I mention these distinctions at the outset to give you some idea of the range of phenomena we will be considering, and to set the stage for our central "problem of representation": what is a mental representation, exactly, and how do we go about deciding whether there are any? We know there are public representations of various kinds: words, maps, and pictures, among others. Are there any representations that are properly thought of as *psychological*?

Minimally, a representation is something that possesses *semantic* properties: a truth value, a satisfaction value (i.e. satisfied or not), truth conditions, satisfaction conditions, reference, or content. So here is one way we might proceed, which we can call "the simple strategy": first, figure out what endows something with semantic properties, and then see if any of the objects that have semantic properties are mental or psychological. The simple strategy might ultimately work, but there are at least two *prima facie* problems with it. First, there is substantial disagreement over what endows an object with semantic properties, and even whether such properties are scientifically accessible at all. Debates about whether thermostats or dogs harbour real representations are often just disputes about whether these things possess states with semantic properties. (The subject of what, if any, natural properties ground semantic properties is the topic of the next chapter.) The second problem with the simple strategy is that there is a danger of talking past one another: not everyone accepts that if something has semantic properties, it must be a species of representation. Many notions of representation require something more. Let us now look at the various notions of representation in common use.

<3>Notions of representation</3>

<4>The degree-of-structure axis</4>

<5>The minimalist or purely semantic notion</5>

The most general notion has already been given: the purely semantic notion, i.e. if it has semantic properties, then it is a representation. On this notion, if folk psychological ascriptions of belief, desire, perception, and intention are true, and such states have semantic properties (which is uncontroversial), then there exist mental representations. Suppose John judges that the sun is bright. On the purely semantic notion, that judgment of John's is a mental representation. Indeed, on the purely semantic notion, mental representation is implied by any true application of a predicate that is both psychological and semantic.

On this minimalist notion, only a radical eliminativist would deny that there are mental representations. It is not clear that there are any such radical eliminativists. The Churchlands are well-known as eliminativists, but they accept that there are mental representations, denying only the existence of the mental representations countenanced by folk psychology (P. S. Churchland 1986; P. M. Churchland 1989). Both Quine (at least in some moods – see his 1960) and Stich (in his 1983) deny semantic properties any scientific credentials. However, Quine is neutral between an eliminativist and a reductionist line, arguing that there is no clear distinction between the two; in more recent work, Stich has made a similar response (Stich 2001). Even in his (1983), Stich seems to accept the existence of mental representations, though maintaining that for scientific purposes they must be individuated syntactically. Why these syntactically individuated types should count as representations is less clear – so the early Stich may come as close as anyone has to denying the existence of mental representations on the purely semantic notion of representation. That said, he acknowledges that, although content-bearing states may play no role in a correct psychological theory, room could be made for them in common-sense, just as there is room in common sense for tables and chairs, although no scientific theory adverts to them.

A purely instrumentalist view, according to which contentful attributions in psychology (whether folk or scientific) are merely a useful fiction, also does not have any clear advocates. Some take Dennett to flirt with such a view (Baker 1995), but in recent times, Dennett has tried to make clear that his "intentional stance" theory does not question the truth of contentful mental state ascriptions (Dennett 1991), which would make him a mental representationalist on the minimalist purely semantic notion.

<5>The thin notion</5>

Let us move now to a somewhat more stringent notion of representation, one that includes fewer theorists under the mental representationalist umbrella. On the purely semantic notion, mental representation requires only the true application of predicates that were both psychological and semantic. On a less minimalist but still rather thin notion of representation, application of a predicate that is both psychological and semantic must be made true by an identifiable state or object harboured by the agent possessing the representation. On this "thin" (as opposed to minimalist, purely semantic) notion, if the fact that John judges that the sun is bright is made true by an identifiable, repeatable neurological state of John, for instance, this neurological state would count as a representation.[1]

The thin notion of representation is strong enough to push some versions of behaviourism out of the representationalist realm (see Chapter 6 of this volume). A Ryleian behaviourist, for example, maintains that mental state attributions are true in virtue of the obtaining of certain counterfactuals concerning behaviour. There need be no identifiable state of the agent that makes them true. So the Ryleian behaviourist denies that beliefs and desires, for instance, need be representations or composed of representations.

Dennett's theory of propositional attitudes also falls on the non-representationalist side of the divide as determined by the thin notion. (It is less clear that the sub-personal-level "representations" Dennett talks about are not representations on the thin notion – see Millikan [2000] and Dennett's response, contained there.) According to Dennett (1987b), semantic, psychological predicates are applied from the "intentional stance," which may legitimately be adopted towards any object whose behaviour is best predicted by assuming that it is trying to get what it wants, in light of what it believes. Such a stance is legitimately adopted when it offers advantages in terms of predictive power over two alternative stances, namely the physical stance and the design stance. (Predictive power is a function of both accuracy and simplicity.) The physical stance predicts by appeal to physical properties and laws; the design stance predicts by assuming intelligent design for a purpose. For instance, consider a simple thermostat based upon a bimetallic strip. Its behaviour may be predicted from the physical stance by appeal to laws concerning the expansion of different metals as well as electric circuits. Its behaviour may also be predicted from the design stance, by assuming that it is designed to keep the temperature at its set point. Its behaviour may further be predicted from the intentional stance, by assuming that it *wants* to keep the temperature at its set point. If it believes that the temperature is higher than its set point, it engages its cooling system because it believes that will bring the temperature back to its set point. In this case, there may be some advantage in predictive power by adopting the design stance over the physical stance: less input information is needed to make the prediction (one doesn't need to know the composition of the lead wires, for instance) and the inferences involved are simpler. However, nothing is gained by adopting the intentional stance rather than the design stance. Neither is the predictive process any simpler, nor is it any more accurate.

Things are otherwise, though, when a dog owner pronounces "Walkies!" within earshot of her Labrador retriever. Adopting the intentional stance allows us to predict that the dog will become excited and perhaps go to fetch a leash. Predicting this behaviour accurately from the physical stance or design stance would require vastly more input information, about the internal structure or design of the dog, and its relation to its environment. (For instance, a separate calculation would have to be made in order to predict that the dog would proceed *around* the table rather than crashing into it, or one would have to know the design details of its navigational mechanisms.) This is a clear case, then, in which the intentional stance may legitimately be adopted.

On Dennett's view, what makes the adoption of the intentional stance legitimate, and therefore what makes the application of semantic, psychological predicates true, are the patterns of behaviour that an object (whether thermostat, dog, or person) exhibits (Dennett 1991). In particular, these patterns must be characterizable in semantic, rational terms. However, as for Ryle (Dennett's intellectual father), there need be no identifiable internal state that is causally responsible for the emergence of this pattern. In principle, every time its owner utters "Walkies!," it could be a *different* condition of the dog that leads to its excited retrieval of its leash, and this would not impugn the attribution of a belief to the dog that it is about to go for a walk, nor its having retrieved the leash for this reason. Thus satisfying Dennett's conditions for legitimate use of the intentional stance is not sufficient for an agent to harbour mental representations, on the thin notion.

There are other theoretical positions that the thin notion deems non-representational. Some dynamical systems theories of cognition claim that there is such a large degree of interdependence among parts of the system, that nothing less than the whole system may be identified as the possessor of semantic properties. On this view, there is no individuable internal state or process than can be separated out from the rest of the system and identified as serving the representational role in question. Any such candidate state or process is causally coupled with so many other states or processes in the system that it would be arbitrary to assign *it* the representational role rather than some other part (Thelen and Smith 1994; Kelso 1995; van Gelder 1995), possibly Brooks, 1991). To the extent that these dynamicists are anti-representational, they are so based on the thin notion of representation. There is a problem with in evaluating such a claim, however, since dynamicists frequently model a system only using abstract systems of mathematical equations, while eschewing any discussion of mechanism (Abrahamsen and Bechtel 2006). Despite their claims to the contrary, it seems reasonable to expect a complete explanation to include details of how these equations are realized in a physical system, and once those details become available it may prove possible to identify physical entities that play representational roles. Dynamicists counter that such representations are explanatorily otiose, and so there is no reason to admit them into a scientific theory of the mind.

This dispute highlights an important division in the sorts of reasons that have been advanced for believing in mental representations. On the one side, there are *explanatory* reasons: do mental representations earn their keep as explanatory postulates? Are they necessary parts of a complete science of the mind (see Stich 1983)? On the other side, there are *ontological* reasons: whatever their explanatory import, given a certain notion of mental representation, are there any such things? It could be, for instance, that the anti-representationalist dynamicists are right about the explanatory question, but the representationalists are right about the ontological question. That said, the two types of reasons are not entirely insulated from one another: one of the best reasons for believing in the existence of something is if it makes a causal contribution to the system of which it is a part. Indeed, on some views, for a property to exist it *must* be causally relevant to something (Alexander's dictum; see Kim 1998). Depending upon how closely causal contributions map onto explanatory contributions, this could tie the two sorts of reasons for believing in representations very closely indeed. Further, often the only reason we have for believing in something is that it is a theoretical postulate (and thus an explanatorily relevant component) of a wellconfirmed scientific theory. If one does not take introspective evidence very seriously, this could also unify the explanatory and ontological reasons for believing in mental representation. These issues take us into the vast literature on mental causation and psychological explanation (see Chapter 8 of this volume), which we will only touch upon here. They also take us deep into fundamental areas of philosophy of science, epistemology, and metaphysics, where we shall tread hardly at all.

Continuing with the thin notion of representation, the claim that nothing less than the whole system may be identified as the possessor of semantic properties comes not only from some dynamicists, but also from a very different quarter. The thin notion of representation also classifies as anti-representationalists those Wittgensteinians who maintain that it is a mistake to speak of semantic properties pertaining to subpersonal states (e.g. Bennett and Hacker 2003). They maintain that "believes that …" and other semantic psychological predicates apply only to whole persons, and never to parts or aspects of persons, defending this position on ordinary language grounds. Thus they would deny that there could be any such internal state of John which is his judgement, and more generally that there could be internal representations. Perhaps their arguments allow that there could be identifiable *person*-level states of John – e.g. beliefs – that were representations, though it is hard to see how that story would go.) This literature has not had much of an impact on psychology and cognitive science, due to scepticism concerning the intuitions revealed through the analysis of ordinary language. Cognitive scientists and naturalistically inclined philosophers doubt that these intuitions have an evidential weight that could challenge any reasonably supported scientific theory. Wittgensteinians retort that since the scientists use ordinary language, their "theories" are literally nonsense, leading to a standoff.

Some cognitive scientists working in the tradition of *embodied cognition* deny that there are mental representations, on the basis of the thin notion, because they deny that there are any identifiable *internal* representations. The brain, body, and environment are so closely coupled that it would be arbitrary to assign semantic properties only to some restricted internal component (this is partly what Rodney Brooks [1991] has in mind; see also Varela et al. [1991].) Such a view is often closely allied with the dynamicist view summarized above. Alternatively, such theorists sometimes insist that there are indeed mental representations, but that they incorporate portions of the environment. (See Chapter 13 of this volume for further details.) This division in embodiment theorists may coincide with their views on whether the individuation strictures of the thin notion can be satisfied by relevant brain-body-environment entities.

Another sort of theory that the thin notion puts into the non-representationalist camp is a social context theory. On this view, what makes the application of a semantic psychological predicate correct is that the agent to whom the predicate is applied is a member of a certain community, subject to certain rules. Usually the community in question is a *linguistic* community. Since being a member of a certain linguistic community and being thereby subject to certain rules is not a matter of harbouring any particular kinds of internal states (even internal states characterized by relations to the environment), a social context theory does not entail that believers harbour representations, on the thin notion. The theory advocated by Robert Brandom (1994) is a good example.

<5>The notion of representation in the representational theory of mind</5>

We now proceed to a yet thicker notion of representation, which characterizes the representational theory of mind proper (RTM) (Fodor 1975; Sterelny 1990). John is not only capable of judging or believing that the sun is bright, he is also capable of hoping that the sun is bright, wanting the sun to be bright, pretending that the sun is bright, fearing that the sun is bright, etc. These mental states differ, but not in their contents. Following Russell (1918), we can say that John can take the “attitudes” of belief, hope, desire, pretence, fear, etc., toward the proposition that *P*. This possibility for mix-and-match of contents with attitudes seems to apply to any proposition *P* that John can entertain. This would be (best?) explained if there existed, in our minds, a representation which meant that-*P*, such that this representation could somehow be “attached” to the different attitudes (or put, as it were, into the “belief-box,” or the “desire box,” etc.; Schiffer 1981). If John’s mind contained a single representation that the sun is bright ready for melding with different attitudes at different times, we would *expect* John to be capable of hoping or wanting the sun to be bright if he is capable of judging that the sun is bright.[2] By contrast, this mix-and-match would be utterly surprising if John’s attitudes had no such representational part or aspect in common. It would be surprising for the same reason that it would be surprising if Billy could produce, on demand from his box, any of the Mr Potato Head combinations, but it turned out that inside the box was not a potato with interchangeable parts, but rather thousands of complete figurines.

The RTM notion of representation, then, is of an identifiable internal state with semantic properties (as on the thin notion), with the additional requirement that this state can mix-and-match with the attitudes. An example of a theory that is non-representational on the RTM notion is a traditional functionalist theory of the attitudes, which treats them as states that may be compositionally unrelated. On David Lewis's theory, for example, the belief that it is raining is characterized *holus-bolus* as the type of physical state that enters into certain causal tendencies with other mental states (e.g. the belief that there is an umbrella in the corner, and a desire to avoid getting wet) as well as perceptual inputs and behaviour (e.g. seeing and getting the umbrella). While the *hope* that it is raining may be characterized by its causal relations with some of the same items as the *belief* that it is raining, by contrast with RTM there is no assumption that the two attitudes literally share a part or aspect corresponding to the representation "that it is raining."

<5>Syntactic structure</5>

To generate an even more restrictive notion of representation, the Mr Potato Head argument may be applied for one more iteration, producing the view that true representations must possess *syntactic structure*. (Sometimes the term "representational theory of mind" is reserved for a theory that takes mental states to exhibit this level of structure, but usually it is meant more broadly, as above.) Just as the observation that John has the capacity to believe that it is raining as well as the capacity to hope that it is raining (etc.) may lead one to postulate the existence of a separable representation "that it is raining," the observation that John may think "the rain is wet," "the rain is cold," "the snow is wet," and "the snow is cold" (etc.) may lead one to postulate the existence of separable representations of rain, snow, wetness, and coldness. These separable representations, on the syntactic structure view, may be structured to produce more complex representations that express complete thoughts – just as words in a language may be structured to produce sentences. The meanings of the complex structure are a function of the meanings of the parts and the syntactic rules for putting the parts together. Such structure is the leading feature of the “language of thought” theory of mental states, found in Sellars (e.g. his 1969) and fully developed by Fodor (1975) (also see Chapter 17 of this volume). (We should also acknowledge that a system of representation that shares features with representational formats that are non-linguistic [e.g. pictures or models] may also exhibit syntactic structure in this minimal sense.) Some authors reserve the term “representation” for psychological states that exhibit some such structure, e.g. Millikan (1984) (though in more recent work she uses the term more liberally).

<4>Related problems: explicit vs. implicit representation</4>

We have now surveyed one axis on which one might draw a line between the representational and the non-representational. Let us call this axis the “degree-of-structure” axis. It goes from zero structure (the minimalist, purely semantic notion of representation) to a high degree of structure (syntactic structure). One of the many “problems of representation” takes place in the degree-of-structure arena. To what extent are our mental representations (if any) structured? The question may be asked of all the various types of candidate mental representations mentioned initially: personal-level representations and sub-personal-level representations; perceptual, cognitive, and action-oriented representations; stored, active, dispositional, and occurrent representations; and finally, conscious and non-conscious representations. Obviously, different answers may be given for the different types. There is a vast amount of literature on such topics; for illustrative purposes, we shall consider one representative issue.

Sometimes a distinction is drawn between a system representing something *explicitly* vs. representing it *implicitly* or *tacitly*. One of the distinctions these terms are used to mark may be drawn on the degree-of-structure axis. (The terms are also used to mark a somewhat related distinction, between a set of representations and the logical consequences of those representations.) As he often does, Dennett gives a nice illustrative case (Dennett 1978; see also Fodor 1985). You might play chess against a computer many times, and come to be familiar with its strategies. You might say, for instance, that it thinks it should get its queen out early. The computer’s program involves a complex evaluation of possible moves; after considering its consequences, each move is given a numerical value, and the move with the highest value is chosen. As it happens, the programmer did not include a line in the program that assigns a higher value to a move that tends to bring the queen out early. Bringing the queen out early is never used as a factor in evaluating moves. However, it turns out that given other lines of the program, the tendency to bring its queen out early emerges naturally; this is indeed an accurate way of summarizing one aspect of the computer's behaviour. But (so the story goes) the computer does not explicitly represent "It is good to bring your queen out early." Only lines in the program (or data structures in memory) are explicitly represented; the computer only implicitly represents that it should get its queen out early.

There are a number of places one might draw this version of the implicit/explicit distinction on our degree-of-structure axis, but it should be clear that this is an appropriate axis on which to draw it. A line in a computer program will have syntactic structure, but the complex state underlying the computer's tendency to bring its queen out early will not. (Perhaps it is representational only on the minimal notion, or only on the thin notion.) A similar sort of distinction could apply to a system that represents in a non-linguistic manner. A model of the solar system (an "orrery") might represent explicitly that Mercury is in such-and-such a position in its orbit, but only represent implicitly the relative speeds of Mercury and Venus. Its representation that Mercury moves more quickly than Venus is a consequence of the dispositional setup of its gears; there is no discrete structure that varies with the relative speeds of the various planets.

The issue with respect to mental representation is whether different types of mental representations are explicit or implicit. (Of course, if one draws the representational–non-representational line at a different position on the degree-of-structure axis, the issue could become one between representing and not representing.) Is a subconscious Freudian desire to kill your father explicit or implicit? Are the rules hypothesized to explain how the visual system arrives at an interpretation of the incoming light explicit or implicit? And even: is your belief that your name is such-and-such explicit or implicit? In a later paper (Dennett 1987a), Dennett remarks that the dispositions built in to the hardware of a common calculator ought to be thought of as representing the truths of arithmetic only implicitly (though he uses the term "tacitly"), and connectionists have sometimes claimed that all representation in connectionist systems is merely implicit. Both of these claims ought to be evaluated empirically by examining the degree of structure that is present in these systems.[3] Similarly for our psychological systems.

<4>The degree-of-systematicity axis: from detectors to speakers</4>

Let us now return to our central issue of what counts as a mental representation. The degree-of-structure axis is only one of many. We shall now briefly examine several other such axes upon which one might draw a line between the representational and the non-representational.

Related to the degree-of-structure axis is the degree-of-systematicity axis. On all views of representation, it seems, in order to represent something, a system must be capable of representing other things as well, in ways that are systematically related to one another. The minimal degree of systematicity required has been described by Ruth Millikan (1984). This minimal degree is such that some possible variation in the representation maps onto some possible variation in the represented. This applies to even the simplest representers, for instance a metal detector. The metal detector beeps when it is brought near metal; on some views this means that it represents the presence or absence of metal here now. The beeping varies in the following ways: presence/absence, time, and place. These variations map onto the presence/absence of metal, the time, and the location.[4] Thus on this very weak account of systematicity, even simple detectors have it. One can imagine even (semantically) simpler detectors that are insensitive to place – they detect the presence of something anywhere in the universe. Perhaps it is even conceptually possible (though surely physically impossible) to have a sort of detector that detects the presence of something anywhere at anytime - e.g. something that detects whether the element with atomic number 369 has existed or will ever exist. Still, such a detector would exhibit a minimal sort of systematicity – it is capable of representing *either* the presence or absence of the element with atomic number 369. Any token representation in such a detector is a member of a (very small) system of possible representations.

Systematic relations are a common requirement for something to count as a mental representation (see e.g. Clark 1997: Ch. 8; Haugeland, 1991). The degree-of-systematicity axis starts with very simple detectors, extends to systems of such detectors, and further runs to complex models with many interrelations (think of a model of terrestrial weather). At the extreme end, we have *extreme representational holism*, which requires the presence of a vast system of representations possessing sophisticated inferential relations, on the order of the sort of system of belief possessed by a normal adult human (see e.g. Davidson 1973).[5] According to the representational holist, you cannot judge that the sun is bright unless you are capable of judging a host of other things, for example that the sun is circular, and therefore that something circular is bright; that the sun is brighter than a match; that the sun is called "the sun," and therefore that something called "the sun" is bright; etc. According to Davidson, there are no *particular* judgements one must be capable of making if one is to judge that the sun is bright, but one must be capable of making a large number of related judgements and acting on them (including expressing them linguistically). Only a holistic representational system like this is capable of being *interpreted* by an observer, and there is no representation without the possibility of interpretation. (You will note the similarity between Davidson's view and Dennett's. One difference is that Dennett does not make such an holistic systematicity criterial of representation, although one might suspect that the advantages of the intentional stance apply exactly to a complex system like this.)

<4>Similarity and iconic representations</4>

Another dimension that is sometimes taken as criterial of representation is *similarity to what is represented*. In medieval philosophy (particularly in Aquinas, following translations of Avicenna and ultimately derived from Aristotle), the term "representation" in psychology is restricted to mental images, which were thought to resemble their objects (Lagerlund 2004). When Aquinas discusses concepts, he calls them *verba* or mental words, and never refers to them with the term "representation." In the modern sense of representation under discussion here, this difference is typically thought to mark a distinction between *types* of representations, rather than between representation and non-representation. That said, some philosophers require that representations resemble what they represent, where the resemblance is usually *relational* resemblance, or isomorphism/homomorphism (Millikan 1984; Gallistel

1990; Swoyer 1991; Cummins 1996).

The debate over whether there are any mental representations that resemble what they represent ("iconic representations") has been a long and heated one. To understand this debate, it is essential to distinguish between a representational *vehicle* and that vehicle's content. A representational vehicle is a (physical) object or event, like a slab of marble, a written word, or a neural firing pattern. (These can be tokens or types.) If that object or event is a representational vehicle, it will also have a meaning or content. The debate under consideration is about representational vehicles, not their contents. A bust of Julius Caesar is an iconic representation if the shaped slab of marble that constitutes it resembles (or is supposed to resemble) Julius Caesar as he was when he was alive. By contrast, the name "Julius Caesar" exhibits no such resemblance, nor is it supposed to – it exhibits an arbitrary structure *vis-à-vis* its object, and is "symbolic" rather than iconic.

A naïve argument in favour of the existence of iconic representations is an introspective one: when one introspects one's mental images, one can see that they resemble their objects. My image of my mother looks just like my mother, one might say. However, while introspection may be a reliable way of registering the *contents* of representations, it is at least highly doubtful that it also registers the intrinsic properties of those representations' vehicles (e.g. the intrinsic properties of neural states). So this introspective argument is little evidence that the representational *vehicles* of mental imagery are iconic. More theoretical arguments are needed, such as those based on the Shepard and Metzler rotation experiments (Shepard and Metzler 1971). In these experiments, subjects were given two pictures of three-dimensional objects that might or might not be rotations of each other, and they had to judge whether they were or not. Shepard and Metzler found that the length of time it took to make the judgement was predicted by a model under which there was a transformation of the representational vehicle corresponding to a rotation of one of the objects to look for a "match." The further the rotation, the longer the judgement took. This suggests that the underlying vehicle is iconic. This sort of argument has also been defended by Stephen Kosslyn (1994), who, in addition, cites brain imaging evidence (i.e. looking at the actual vehicles). The iconic view is opposed most prominently by Zenon Pylyshyn (see e.g. his 2003 and 1999).

<4>Use: causal role and function</4>

We now turn to a series of distinctions related to how an internal state is *used* in the psychological economy. It is quite plausible that, for instance, representation-hood cannot be determined on the basis of degree of structure or degree of similarity alone. For example, the building-hating but sky-loving city dweller might trim his hedge to block out, very precisely, the downtown skyline, and his hedge would then resemble the downtown buildings. But it is far from obvious that the hedge thereby represents the downtown buildings. If, however, someone were to make *use* of the similarity to choose their afternoon destination (the tallest skyscraper), then the representational status of the hedge becomes more plausible. It is not a map of the downtown skyline unless it is *used* as such. Similarly, the map-like organization of a bunch of neurons in somatosensory cortex does not necessarily represent the relative positions of the limbs unless that mapping can be *used* by higher areas, so that it has functional import (see also Chapter 23 of this volume).

There are two main ways of understanding use naturalistically, as causal role or as normative or teleological functional role (corresponding to two broad types of functionalism – see Chapter 10 of this volume). For a representation to occupy a particular type of (pure) *causal role* is for it to be located in a particular pattern of causes and effects (which might extend many causal steps upstream and/or downstream). (Usually the requirement will be for the representation to be *disposed* to enter into a particular causal pattern, rather than for it actually to be occupying it right now.) Typically a *teleological* functional role is also a causal role, but it is one that the representation is *supposed* to enter into; that is its job or function. In artificial systems, that job is usually assigned by a designer. In natural systems (e.g. human cognitive systems), that job is alleged to be assigned by evolution and/or learning (see the next chapter for details). Any of the bases for dividing representations from non-representations that are related to use might be given a pure causal or, alternatively, a teleological reading.

<5>Representation hungriness</5>

We can use a nice term from Andy Clark (1997) to name our first use-based axis often suggested to mark the representational/non-representational divide: the degree of "representation hungriness" of the task for which a putative representation is deployed. Following Haugeland (1991), Clark suggests that one type of representation-hungry problem arises when a system "must coordinate its behaviors with environmental features that are not always reliably present to the system" (1997: 167). Some examples would be having the capacity to plan a route ahead of time, or to recall my distant lair, or to imagine what would happen if I tipped this rock off the cliff onto Wily Coyote. This idea of representation is strongly connected to the notion of a representation as a stand-in for an absent feature or state of affairs. The intuition here is that in the absence of what a representation is about, there is a need for some sort of internal "stand-in" to have a psychological effect. My distant lair cannot *directly* affect my thinking or behaviour if it is not present to me, so if my lair appears to affect my thinking or behaviour it must do so via some intermediary. (A real example from the psychological literature: C. R. Gallistel [1990] argues that ants have a representation of their location with respect to their lair, since, in a featureless desert, they can walk the correct distance in a straight line no matter what their previous path was.) By contrast, when a flower adjusts its position to keep constantly facing the sun, no "stand-in" is needed. The sun can directly cause the change in the sunflower's position.

The other type of representation-hungry problem, according to Clark, "involves selective sensitivity to states of affairs whose physical manifestations are complex and unruly" (1997: 167). If someone has the capacity to respond to or reason about events that are characterized by moral turpitude, for instance, it is reasonable to expect their cognitive system to be set up in such a way that "all the various superficially different inputs are first assimilated to a common inner state or process such that further processing (reasoning) can be defined over the inner correlate." Events that are extremely dissimilar physically may all be perceived by a cognitive system as infected by moral turpitude. Treating them the same in this way would seem to require an internal stand-in for moral turpitude. By the same token, two very similar events, physically speaking, might *fail* to share the characteristic of moral turpitude (if one act is committed by a large but young child, for instance). This time the system must be able to treat very similar perceptual encodings as calling for entirely different responses, and again, an internal state seems necessary "to guide behavior despite the effective unfriendliness of the ambient environmental signal" (168).[6]

Clark does not exactly present it this way, but involvement in representation-hungry problems is sometimes taken to be criterial of genuine representation. Following Gibson (1979), ecological psychologists deny that there are internal representations mediating many psychological processes precisely because they maintain there are few if any representation-hungry problems (and so no need for an internal "stand-in").[7] They claim that there is always some invariant feature of the environment that is available to be perceived and to drive behaviour directly, as long as the system is sensitive to that feature. If this seems implausible, consider that some of the features that can be directly perceived, according to ecological psychologists, are certain dispositional or relational features called "affordances" – so that I can directly perceive that the rock on the cliff "affords" the squishing of Wily Coyote. To take a simpler example, consider when I look at some steps, and judge whether they are climbable (in the normal way) or not. It turns out that there is information directly available to me (step frequency, riser height, stair diagonal, leg length, strength and flexibility, and body weight) that fully specifies whether those steps are climbable for me (Konczak et al. 1992; Warren 1984). I need not *imagine* myself climbing the stairs to discover this (thus making use of an internal stand-in); I can simply register the combination of specifying information. (Hubert Dreyfus [2002] seems to have something similar in mind when he critiques representationalism, and it is also part of what Rodney Brooks [1991] contends.)

So judging climbability turns out *not* to be a representation-hungry problem, in Clark's sense, at least not for the reason of an *absent* (as opposed to unruly) environmental signal.[8] Whether more complicated cases, like the Wily Coyote case, or spatial navigation, will also turn out not to be representation-hungry is something that the ecological psychologists have yet to demonstrate to sceptical mainstream cognitive scientists. And since ecological psychologists in general eschew talk of mechanisms, they have not provided much reason to doubt that internal stand-ins are sometimes necessary in order to deal with an *unruly* environmental signal, Clark's second type of representation-hungry problem.

<5>Other suggested typifying roles</5>

What exactly *is* the typifying causal role or teleological function of representation? As we have just seen, disagreements may arise as to an item's representational status based on a fundamental disagreement about what sort of functional role makes a representation a representation.[9] The ecological psychologists seem to assume that representation occurs only when the cognitive system is not driven by some ambient signal in the environment, whereas mainstream cognitive scientists allow that representation may occur when the system responds to an environmental signal that is present but unruly. A number of other typifying functional roles of representation have been implied or suggested in the literature.

One obvious one is storage in memory. Information about the environment is collected, but not lost – it is kept for later use. This is one major role often envisioned for representation, and it is clearly linked to Clark's first representation-hungry problem (sensitivity to absent stimuli). Another major role commonly envisioned for representation is information processing – once stored, the information may be manipulated in combination with other pieces of information to generate new representations. A third major role is in problem solving, especially problem solving that involves hypothetical reasoning (e.g. the Tower of Hanoi problem). For example, Soar is a cognitive architecture from artificial intelligence that is primarily designed for problem-solving, and its leading characteristic is the manipulation of representations to explore (in "working memory") the various options in a "problem space," eventually settling on a procedure that will accomplish the task it has been set (Rosenbloom et al. 1992). (This is another instance of introducing representations to address Clark's first representation-hungry problem.) However, sometimes much simpler functional roles are taken to be criterial of representation. In the neuroscientific literature, the term "representation" is often used just in case there is a neural "detector" of some type of environmental stimulus, which need not be particularly unruly. (For instance, ganglion cells in the retina are sometime said to "represent" the impingement of multiple photons in a single location.) Usually it is also (implicitly) required that the signal thus generated can be used in further processing, though not necessarily computational processing in any robust sense.

In the psychological literature, representations (especially concepts) are often assumed to have the role of "classifiers," with the role of assimilating a number of (possibly unruly) environmental signals. These classifiers then may play a number of other roles, including, for example: (1) Economy of storage: Rather than remembering each experience individually, a cognitive system may remember features as pertaining to a small number of situation types (thus “dogs are furry” rather than “Rover is furry, Spot is furry, Snoopy is furry…”). (2) Economy of interface: This proposed function is closely related to the handling of “unruly “ environmental signals, as above. Perceptual inputs need to be organized to facilitate action, and this can require the involvement of simplifying intermediaries. The sort of reasoning implemented by a classical “symbol manipulation” cognitive architecture (see Chapter 7 of this volume) can be seen as this kind of action-facilitating interface. (3) Communication: One of the actions facilitated by representations understood as classifiers is communication. Language is a small bandwidth system, so the economy introduced via classification also helps enable communication. (4) Identification of kinds and induction: If the features unified by a single representation are features of a real kind (e.g. dogs rather than the arbitrary class of red balloons, cups, and asparagus), the representation will be useful for inductive inference (if one dog is furry and brown, it is more likely that another dog is furry and brown; whereas the pointiness and greenness of asparagus does not make it more likely that a red balloon will be green and pointy). It will also be useful for organizing action, since items that share fundamental similarities are more likely to be appropriately subjected to similar actions.

<4>Combination approaches</4>

As one might expect, researchers often make use of multiple criteria simultaneously in wielding the term “representation.” This is particularly true in the scientific literature, and often the criteria being used are merely implicit. Gallistel and Gibbon (2001) provide a nice exception, which I will present as an example of the combination approach. In arguing for their view that classical (Pavlovian) conditioning is a representational process, they contrast a computational/representational process with a process of simple association. They rely on a multi-pronged account of what makes something a representation. First, it must have a semantic content, or as they put it, it must “encode information about objectively specifiable properties of the conditioning experience.” If classical conditioning were implemented by an associative process, there would be no representational separation of the variables that can affect the strengthening and weakening of the association between the conditioned and unconditioned stimuli (in Pavlov's famous experiments, CS = bell, US = food). There are many such variables, including (for example) the length of time between trials, the length of time between reinforcements, the number of reinforcements, and the delay in reinforcement. Because all of these variables would be confounded together in one association (a process of strengthening or weakening a connection), Gallistel and Gibbon maintain there is no determinate semantic content (about time, or number of trials, etc.) that could be attached to that associative process, or aspects of it. Since it would not have a determinate semantics, then, such a process could not be representational. They go on to argue that the experimental results suggest that what is going on is not in fact associative, but rather involves real representational separation. In particular, conditioning is insensitive to the time scale of the various intervals in the experimental protocol. As long as the ratios between the length of time between reinforcements, reinforcement delay, etc., are kept constant, the same effects on reinforcement strength are observed. The best explanation for this, they argue, is that the system keeps *separate* track of the various time intervals involved. This separation allows for determinate semantics and thus genuine representation.

The second part of their combination requirement for representation is a functional one: storage in memory. This is so that the third part may be satisfied: subsequent manipulation to generate new representations ("information processing"). In their representational account of conditioning, the various time intervals involved are "recorded in memory for later use in the computation of the decision variables on which conditioned responding is based." By contrast, "the associative bond does not participate in information processing (computational) operations. Associations, unlike symbols, are not added, subtracted, multiplied, and divided in order to generate new associations."

So here we have a nice example of the combination approach. It is also a nice example of when explanatory value grounds the postulation of representations. In Gallistel and Gibbon's model of conditioning, the question they try to answer is this: how does the animal manage to display time-scale invariance in its learning? They argue that it requires (1) variable separation; (2) variable storage; and (3) variable manipulation – and it is those related explanatory needs that lead them to postulate the existence of representations in the conditioning process (given their criteria for representation-hood). By contrast, if an associative model were explanatorily sufficient, no representations would need to be postulated.

<3>Conclusion: other problems of representation</3>

This review has been organized around the central strategy of figuring out what makes something count as a representation, and from there deciding what things count as *mental* representations. We have seen a wide variety of proposed criteria for representationhood, including degree of structure, degree of systematicity, similarity/isomorphism, representation hungriness, playing a role in detection, in storage, in information processing, in problem solving, or in classification. Many debates about the nature and role of various kinds of mental representations can be illuminated by paying careful attention to the participants' assumptions about what makes something a representation. We saw a few examples, but we have neglected many others. Fortunately most of these are treated elsewhere in this volume.

For example, there is controversy over the extent to which the outside world is in some sense a part of mental representation. Some advocates of embodied cognition claim that the outside world is literally a part of our mental representations (see Chapter 13 of this volume), and content externalists argue that a representation's content – and therefore the representation's identity – depends on things external to the mind (see the next chapter). There have also been and continue to be extensive debates about the format and nature of the various types of mental representations: are they classical, connectionist, or dynamical (Chapters 7, 12, and 28)? Are they digital or analog (Chapter 13)? Is there a language of thought, and is that language *sui generis* or is it a public language (Chapter 17)? If there are mental modules, do these modules share a *lingua franca*, or each "speak" a different language (Chapter 18)? Are there stored mental "rules" that dictate psychological processing (Chapter 12)? Are there both conceptual and non-conceptual representations (Gunther 2003)? Are there non-representational mental states, e.g. are sensory states non-representational (Chapters 29 and 35)? Are all mental representations consciously accessible (Chapter 31)? Are there innate mental representations (Chapter 19)? All of these other problems, though, will require an answer to the central question of what counts as a mental representation.

<3>Notes</3>

[1] This requirement is imposed by, among others, John Haugeland (1991) and Andy Clark (1997).

[2] Or alternatively, if something that could produce invariant that-the-sun-is-bright representations on demand.

[3] It is precisely the lack of structure in connectionist systems that forms the basis for Fodor and Pylyshyn's famous attack on them – see Sharkey and Sharkey (this volume).

[4] Millikan calls this corresponding variation an "isomorphism," though it should be understood that the isomorphism is not between a set of currently active representations and a set of circumstances in the world; rather it is an isomorphism between a set of possible currently active representations (beeps at different times and places) and a set of possible circumstances – the isomorphism could be called a "modal isomorphism." This contrasts with a picture, or more generally an "iconic" representation, in which the structure of a single representation maps onto the structure of some bit of the environment. This sort of isomorphism shall be considered below.

[5] Although they often go hand-in-hand, representational holism, according to which representational status depends on holistic systematic relations, should be distinguished from content holism, according to which a representation's content depends upon these relations. One could be a representational holist, and still accept Fodor and Lepore's (1992) central argument against *content* holism, to the effect that content holism is incompatible with a compositional syntactic structure.

[6] As Clark notes, the classical computational model of vision has developed the view that even ordinary cases of perception, like perceiving distance, in fact involve dealing with similarly unfriendly environmental signals, thus their invocation of a multitude of internal representations.

[7] Sometimes ecological psychologists talk as though representation hungriness is *criterial* for genuine representation, but other times they seem to take it as

merely *evidence* for it, i.e. an internal “stand-in.”

[8] Gibson also makes the point that the registration of an environmental signal takes time – if that time can be indefinitely long, we may even obviate the need for memory, if memory is to be understood as the storage of representations. Instead, we may think of the cognitive system as registering the presence of information in the environment over the course of seconds, hours, days, or even years.

[9] I’ll now use this term, “functional role,” generically to include both causal role and teleological role readings.

<3>References</3>

Abrahamsen, A., and Bechtel, W. (2006) “Phenomena and Mechanisms: Putting the Symbolic, Connectionist, and Dynamical Systems Debate in Broader Perspective,” in Stainton, R. J. (ed.), *Contemporary Debates in Cognitive Science,* Oxford: Blackwell, pp. 159–86.

Baker, L. R. (1995) *Explaining Attitudes: A Practical Approach to the Mind*, Cambridge: Cambridge University Press.

Bennett, M. R., and Hacker, P. M. S. (2003) *Philosophical Foundations of Neuroscience*, Malden, MA: Blackwell.

Brandom, R. (1994) *Making It Explicit*, Cambridge, MA: Harvard University Press.

Brooks, R. (1991) Intelligence without Representation,” *Artificial Intelligence* 47: 139–59.

Churchland, P. M. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press.

Churchland, P. S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, Cambridge, MA: MIT Press.

Clark, A. (1997) *Being There*, Cambridge, MA: MIT Press.

Cummins, R. (1996) *Representations, Targets, and Attitudes*, Cambridge, MA: MIT Press.

Davidson, D. (1973) “Radical Interpretation,” *Dialectica* 27: 314–28.

Dennett, D. (1978) “A Cure for the Common Code,” in *Brainstorms,* Cambridge, MA: MIT Press, pp. 90–108.

——— (1987a) “Styles of Mental Representation,” in *The Intentional Stance,* Cambridge, MA: MIT Press, pp. 213–25.

——— (1991) “Real Patterns,” *Journal of Philosophy* 88, no. 1: 27–51.

——— (1987b) *The Intentional Stance*, Cambridge, MA: MIT Press.

Dreyfus, H. (2002) “Intelligence without representation,” *Phenomenology and the Cognitive Sciences* 1, no. 4: 367–83.

Fodor, J. A. (1985) “Fodor’s Guide to Mental Representation: The Intelligent Auntie’s Vade-Mecum,” *Mind* 94: 55–97.

——— (1975) *The Language of Thought*, New York: Thomas Y. Crowell & Co.

Fodor, J., and Lepore, E. (1992) *Holism: A Shopper’s Guide*, Oxford: Blackwell.

Gallistel, C. R. (1990) *The Organization of Learning*, Cambridge, MA: MIT Press.

Gallistel, C. R., and Gibbon, J. (2001) “Computational vs. Associative Models of Simple Conditioning,” *Current Directions in Psychological Science* 10: 146–50.

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception*, Boston: Houghton-Mifflin.

Gunther, Y. H. (ed.) (2003) *Essays on Nonconceptual Content*, Cambridge, MA: MIT Press.

Haugeland, J. (1991) “Representational Genera,” in W. Ramsey, S. Stich, and D. E. Rumelhart (eds), *Philosophy and Connectionist Theory*, Hillsdale, NJ: Erlbaum, pp. 61–89.

Kelso, J. A. S. (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior*, Cambridge, MA: MIT Press.

Kim, J. (1998) *Mind in a Physical World*, Cambridge, MA: MIT Press.

Konczak, J., Meeuwsen, H. J., and Cress, M. E. (1992) “Changing Affordances in Stair Climbing: The Perception of Maximum Climbability in Young and Older Adults,” *Journal of Experimental Psychology: Human Perception and Performance* 18, no. 3: 691–97.

Kosslyn, S. M. (1994) *Image and Brain*, Cambridge, MA: MIT Press.

Lagerlund, H. (2004) “Mental Representation in Medieval Philosophy”; available: http://plato.stanford.edu/archives/sum2004/entries/representation-medieval/

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: W. H. Freeman.

Millikan, R. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.

——— (2000) "Reading Mother Nature's Mind," in D. Ross, A. Brook, and D. Thompson (eds), *Dennett's Philosophy: A Comprehensive Assessment*, Cambridge, Mass.: MIT Press, pp. 55–76.

Pylyshyn, Z. W. (2003) Return of the Mental Image: Are There Really Pictures in the Brain?" *Trends in Cognitive Sciences* 7(3): 113–18.

——— (1999) "Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception," *Behavioral and Brain Sciences* 22, no. 3: 341–423.

Quine, W. V. O. (1960) *Word and Object*, Cambridge, MA: MIT Press.

Rosenbloom, P. S., Laird, J. E., and Newell, A. (1992) *The Soar Papers: Research on Integrated Intelligence*, Cambridge, MA: MIT Press.

Russell, B. (1918) "The Philosophy of Logical Atomism," *Monist* 28: 495–527.

Schiffer, S. (1981) "Truth and the Theory of Content," in H. Parret and J. Bouveresse (eds), *Meaning and Understanding*, New York: De Gruyter, pp. 204–22.

Sellars, W. (1969) "Language as Thought and as Communication," *Philosophy and Phenomenological Research* 29: 506–27.

Shepard, R. N., and Metzler, J. (1971) "Mental Rotation of Three-Dimensional Objects," *Science* 171: 701–3.

Sterelny, K. (1990) *The Representational Theory of Mind*, Oxford: Blackwell.

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

——— (2001) *Deconstructing the Mind*, New York: Oxford University Press.

Swoyer, C. (1991) "Structural Representation and Surrogative Reasoning" *Synthese* 87: 449–508.

Thelen, E., and Smith, L. B. (1994) *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge, MA: MIT Press.

van Gelder, T. (1995) "What Might Cognition Be, If Not Computation?" *Journal of Philosophy* 91: 345–81.

Varela, F., Thompson, E., and Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.

Warren, W. H. (1984) "Perceiving Affordances: Visual Guidance of Stair Climbing," *Journal of Experimental Psychology: Human Perception and Performance* 10: 683–703.