Intention

(Doing Away with Mental Representation)

Jay L Garfield

Department of Philosophy, Smith College

Department of Philosophy, University of Melbourne

Central Institute of Higher Tibetan Studies

1.   The Metaphor

Wilfrid Sellars used to be fond of saying two things about metaphors in philosophy: "Philosophy is nothing but the construction of metaphors. We pile them higher and higher until, like a house of cards, the entire structure collapses. Then we search for a new set of metaphors and start the process all over again." But also, "all metaphors limp. A metaphor always needs a commentary, and in the end it is as likely to mislead as it is to illuminate."

While the thoughts I express in this paper are continuous with those I have been expressing for some time, the immediate impetus derived both from a re-thinking of the importance of recent work in embodied cognition—particularly that of Clark and Haugeland, some conversations with Terry Dartnall, and a re-thinking of the importance of the work and remarks of a few of those

figures in the history of philosophy to whom I regularly return: Berkeley, Hume, Kant and Wittgenstein.

Mental representation is a metaphor. It has perhaps become so entrenched that it appears to have been frozen, and it is easy to lose sight of its metaphorical character. *Literally,* a representation is a re-presentation, a symbol that stands for something else because that thing can't be with us. I send my parents photos of the grandchildren because e-mail is cheaper than air tickets. I consult a map of Adelaide to find the shortest route to the philosophy department because wandering through the streets would take too much time. *Perhaps* in a similar sense I use words in this discussion because the very ideas in my head have no way of being transferred directly to yours. It is tempting to think that when you hear them thoughts in your head substitute in further processing for the sounds I produce or for the words they encode. But if I am right in what follows, we should resist this temptation.

When we talk as though beyond photographic, stylized pictorial representations, there are also linguistic representations we make an interesting, and perhaps not entirely literal extension of the concept; when we go one step further and take it that thought demands *mental* representations we continue to borrow the idea of "standing for" and suggest that the vehicles of our thought stand for their referents in much the same way that pictures stand for theirs. This way of talking has, as I have noted, become so terribly entrenched that it appears to encode an obvious truth. Many of us in the philosophy of mind and in cognitive science have become so accustomed to thinking through this metaphor that to question it smacks of eliminativism about the mental, behaviorism, or some equally dark pathology. The only real questions, at least the only questions that most of us debate seriously, then turn out to be those about

the precise mechanism, structure, or other features of our representations. (In fact, even self-avowed eliminativists such as Paul Churchland (2002) use the term without any hesitation, and talk both about representation in the mind and in the brain, reserving their eliminativism for specific *forms* of representations, such as propositional attitudes. I remain a friend of belief, though I find myself becoming a foe of representation, and so position myself at the same time closer to the mainstream and further on the lunatic fringe. Such is the conceptual geography of this field.)

So now I want to disturb this venerable house of cards, and to suggest that this metaphor has had its day; that it no longer even limps, and that it is time to put it out of its misery.

First, a bit of loosening up: As I noted above, we use representations when the original just isn't available. It would be great to see the grandkids, but a picture will have to do. On the other hand, while it would be nice for the grandkids to be among us, would it *really* be so great to get the grandkids into our *skulls*? And what would it even *mean* to get them *into* our *minds*? (Note here that on a properly Humean understanding according to which the mind and its perceptions do not contrast with, but "spread over" the world, this last idea might not be so bad, but in *that* sense we are offering an *alternative* to a representational view of perception and thought.) Do our *thoughts* about them serve as second-best substitutes in the same sense that our photographs do? That is, if my mom asks how my son looks, could I tell her just to think about him and find out for herself? Hardly.

Things are even worse in the case of perception: If I see the children in front of me, they are already *present.*   Why *re-present* them*?* What good would it do me to see them by seeing an image of them? To *what* to which the children are not *already* present would that re-presentation be *present*? And why is that *second* presence of any special value? Finally, must it be re-presented once again in order to be of any use? Surely the regress has to be cut off somewhere and the obvious place to cut if off is at the root. This is not, however, to say that it is obvious just how to cut that regress without loss of explanatory power. That is the burden of this essay.

In a short while I will start talking about how to think about thought without thinking about representation, and how to do so in a robustly realist, intentionalist spirit. But first, having emphasised the metaphorical nature of the idea of representation, I want to prosecute the claim that it is no harmless metaphor. It really does *mislead.*

2.   The Prevalence and Costs of the Metaphor

If I were to ask one hundred contemporary philosophers of mind for the clearest example of an eliminativist with respect to the mental, I'll bet that ninety would answer, within about one hundred milliseconds on average, "Paul Churchland." So it is instructive to turn to one of Churchland's recent discussions of the topic of mental representation, his APA Presidential address:

> Suppose also the internal; character of each of the representational spaces is not *fixed* by some prior decree, either divine or genetic, but is slowly shaped or sculpted by the extended experience of the developing animal, to reflect the peculiar learning procedures embodied in the brain's ongoing business of synaptic modification.....
> ...

So we begin by expanding the number of representational spaces, into the hundreds and thousands, far beyond the Kantians pair. We locate them in discrete anatomical parts of the brain. We make each one of them plastic and multipotent in its semantic content and its conceptual organization. And we reach out to include motor cognition and practical skills, along with perceptual apprehensions and theoretical judgment, as equal partners in our account of human knowledge. [Churchland 2003, p 206]

Whatever it is that Churchland proposes we eliminate in cognitive science, it is clear that it is *not* representations. In fact he supposes here quite blithely that the neurocognitive processes that mediate not only conceptual thought and perception but also motor control and skill acquisition are representational. His brief in an account of a plastic, distributed and superpositional model of representation is, if anything, not the *elimination* of the category of representation from cognitive science, but the *extension* of the category to virtually everything that the nervous system does. This push for homogeneity in theoretical approach may have something to it, though I think that Churchland goes too far. I will, however, take sharp issue with the idea that the relevant homogenizing category ought to be that of representation.

Andy Clark is another who might be cited as an opponent of much explicitly representational accounts of cognition. I will pay closer attention to his work below. But for now note that he shares this presupposition that even the most distributed neural processing, so long as it is teleologically connected to an object, is representational:

A distributed representation is an inner encoding in which the target content is not carried by an individual recourse (e.g. a single neuron) and is not necessarily carried by a spatially localized group of units or neutrons. Instead, the context (concerning e.g., the notion of an individual digit) is carried by a pattern of activationwhich is spread across a population of neurons or units. Distributed encoding present a number of advantages and opportunities, For example, the pattern itself can encode significant structural information in such a way that minor variations in the pattern reflect small but sometimes important differences in what is currently represented. ... [T]he point for now is simply that the brain

may be using quite complex, overlapping, spatially distributed representational schemes even in cases where we might intuitively have expected a simple, spatially localized encoding strategy, as in the case with there M1 motor area. Nature's way, it seems, is to use spatially overlapping distributed encoding to govern related (but nonidentical) types of finger movement. The final picture is thus one in which specific cortical neurons play a role in controlling *several* finger muscles and do so by participating in widely spatially extended patterns of activity which correspond to different types and directions of finger movement. [Clark 1997, pp 132-133]

It is worth noting two features of this brief discussion in getting a fix on just how embedded the representational metaphor is in our thinking about the mind: First, there is the assumption that any kind of information-bearing states or processes are representational, however much they might not look to be—that representation is the default assumption regarding the nature of cognitive activity; second, even the processes that subserve motor control are to be thought of as representational—the drive to treat all cognition homogeneously is not taken as a drive to treat conceptual activity as similar to motor control in virtue of being *non-representatnional,* but the reverse.

We could go on multiplying examples, but I trust the point is made. Even the most radical opponents of "classical," linguistic, or digital models of thought presuppose a representational model. But once we begin talking, thinking and theorising in terms of representation, we find ourselves drawn inexorably into a labyrinth of dialectical dead-ends. They constitute a depressing proportion of recent literature in the foundations of cognitive science, and include such debates as those concerning whether our representations are digital or analog; whether they are imagistic or propositional; whether we share representations with infralingual children and animals; and whether, in pursuing cognitive science, we should be realistic or instrumentalist about these mental stand-ins for reality. In each case we seem to fall into a state of collective antinomy.

In each case, I suspect the reason for the antinomy is that the debate in question proceeds by means of metonymic equivocation: we are led by the representational metaphor to shift conceptual focus from the *contents* of our thoughts (about which many of these questions can be intelligibly and even fruitfully raised) to the *nature* of the thoughts themselves, where the questions and the debates lose their grip.

Debates about whether human representation is digital or analogue have been notably sterile. Reasons for each alternative are matched by rebuttals and equally plausible but equally refuted reasons for the opposing conclusion. There are four possible reasons for this state of affairs: It might just be too hard a question. OR, it might be that some human cognition is analogue (image rotation, say) and some is digital (syntactic analysis). OR it might be, that like the painting, it is digital with respect to some level of analysis or explanatory interest, but analog with respect to another. OR finally, it might be, that like taking hold of the pipe, it doesn't make any sense to ask of thought whether it is one or the other, and that we are really asking questions about the *objects* of thought and confusing them for questions about the *vehicle*.

Detailed argument against each of the first three alternatives would be a long and difficult matter. Here are a few motivators for rejecting each of them, however: questions about whether a particular form of representation is analog or digital simply aren't in general *too hard*. Now, to be sure, questions about the mind are often very hard, but in cognitive science we have become used to real progress on empirical questions of all kinds, and this seems an empirical question. There simply is no independent reason to suspect its enormous difficulty *unless* one accepts

antecedently the representational metaphor of the mental *and* thinks that those representations are especially occult.

The some-some response gets its motivation from intuitions about the phenomenology of mental imagery, apparently corroborated by such data as those of Shepard and Metzler (xxx) on the one hand and about the conviction that the computational model of mind must be at least generally right, on the other. But each of these is deeply problematic: To the extent that our intuitions about what goes on in our mind are any guide to what actually occurs (and they are notably fallible in this regard across the board, as we all know) they are useful with regard to the *contents* of our mental states and processes—that towards which they are directed, and not with regard to the nature of those states and processes themselves. It might well be that we can discover by introspection that we think *about* images, or *about* sentences, but to argue that we have some kind of privileged access to the *vehicles* by means of which we think about those objects .

Nobody doubts this. The rotation studies are designed precisely to get behind mere intuition, and the fact that response time for matching block patterns to targets is directly proportional to the angle of rotation of the pattern with respect to its target suggests that an inner image—an analogue representation—is being manipulated by a process of mental rotation—an analog process defined over analog representations. But this interpretation of those data already presupposes a representational model of thought. Let me explain: If I rotate a physical arrangement of blocks in my hand, the blocks are rotated, not represented. If I watch those blocks being rotated, I must visually detect those blocks, but there is no need to represent them; I can simply see them. There may well be something analog about the seeing: the states in my visual system that respond to the

light reflected off the blocks may vary continuously with their rotation. There may also be some digital features of the seeing. But none of these are features of representations, only of states that mediate detection and interaction. Now, when I *imagine* the blocks being rotated, my cognitive activity is similarly directed on the block, and in fact recruits many of the same brain processes that subserve actual vision. What is going on is a kind of ersatz vision. To the extent that anything is being re-presented it is not the blocks, but the visual or visuo-motor processes involved in seeing them being rotated or in rotating them. Imagining on this account is exactly as representational or as non-representational as vision. Analog processing does not demand analog representation.

The view that thought is in general a set of digital operations defined over a set of representations in a language of thought was once central to our field, but I take it that by now it has had its day. I will not flog that particular dead horse, but only recall some of the blows responsible for its demise: the failure to discover anything in the brain plausible like such symbols; the apparent causal necessity of natural language acquisition for so much human cognition; the success of distributed models of processing as models of cognitive activities, etc... So much for the some-some response. I conclude that it might be a good idea to recuse from the debate about the nature of mental representation and ask instead whether we can do better than representation as a metaphor for the mental.

3. Clarke and Haugeland on Representational Realism

As I said at the outset, a lot of my thinking about this problem was occasioned by another look at recent work on embodied cognition. So now I want to pay attention to an extended interchange between Andy Clark and John Haugeland. Here is Clarke on Haugeland:

Cognitive scientists often talk of both brains and computer models as housing internal representation." This basic idea provided common ground even between the otherwise opposing camps of connectionism and classical artificial intelligence. The differences between connectionists and the classicists concerned only the precise nature of the internal representational system, not its very existence. Classicists believed in a "chunky symbolic," inner economy in which mental contents were tokened as strings of symbols that could be read, copied, and moved by some kind of inner central processing unit.

For all that, explicit, chunky symbolic representations and distributed vectorial connectionist representations were both seen as species of internal representation, properly so called. This overarching species is present, it has been argued, whenever a system meets certain intuitive requirements. Haugeland (1991) unpacks these by describing a system as representation using just in case:

> 4. It must coordinate its behaviours with environmental features that are not always "reliably present to the system."
> 5. It copes with such cases by having something else (in place of A signal directly received from the environment) "stand in" and guide behaviour in its stead.
> 6. That "something else" is part of a more general representational scheme that allows the standing in to occur systematically and allows for a variety of related representational states. (62)

[Clarke 1997, pp 143-144]

Again, we have the presupposition that in talking about inner cognitive states and processes we are talking about representations and operations on them. Clarke notes that Haugeland analyses representations derivatively, beginning with a characterization of the organisms or systems that use representations. For Haugeland, as Clark notes, the crucial characteristic of such systems or organisms is that they coordinate their activity with the (at least sometimes) absent or non-existent, and that they do so by means of a "stand in" that substitutes for that which is not present. That stand in, if it is part of an articulated system of such stand-ins, and processes defined over them, is a representation. Clarke quarrels with the requirement that representations represent the non-present. He considers the neural states that track rats' head positions. These are activation vectors that co-vary with the orientation of the head in the environment :

It seems reasonably clear that by glossing states of the neuronal populations as codings for specific head positions we gain useful explanatory leverage. Such glosses help us understand the flow of information within the system when, for example, we find other neuronal groups (such as

motor control populations) that consume the information encoded in the target population. The strict application of Haugeland's criteria would, however rule out the description of any such inner systems of non –decouplable inner states as genuinely representational. This seems unappealing in virtue of the very real explanatory leverage that the representational gloss provides…. (Clarke 1997, p 145)

A couple of things deserve mention here before we take stock and move to some of Haugeland's more recent remarks: Clarke argues that states such as those tracking and controlling rats' head positions are representational despite the fact that they cannot be decoupled from their putative representational content because they encode information that is used elsewhere in the system. This may seem a small terminological quibble, but it in fact raises an important issue. Clarke is again drawing attention to a certain kind of homogeneity: The processes that track that which is present are not all that different from those that track that which is absent. So to the extent that we take representational status seriously, one set of processes and states is representational exactly to the extent that the other is. I think that intuition is right on target, as is the observation about the character of the cognitive processes in question, both with respect to their intrinsic character and with respect to their relation to their objects. But Haugeland suggests that we only have representation when we have something that can "stand in for" or "re-present" that which is absent. That is absolutely right as well. But these two premises give us, instead of the conclusion that Clarke's rat neuron populations are representational the conclusion that most of the cases that he, Haugeland and others in this field regard as representational are not.

This is not the conclusion to which Haugeland comes, of course. Consider this discussion in the context of a discussion of Cummins' account of representation in Hugh Clapin's excellent volume on the philosophy of mental representation: Discussing a dog's cognitive engagement with a ball, Haugeland says:

...Shelia doesn't have any intenders the function of which is to target *specifically tennis-balls.* Nor, likewise, can any of her representations have specifically them as their contents. Of course, if she has representations, then they must have contents; and then we face the familiar muddle about how to specify them, given that *all* our words are way too laden and sophisticated. But that doesn't strike me as a killer objection so much as an annoying limitation. (Indeed, I think that we ourselves have and use representation that we couldn't put into words). So, taking the scare quotes suitably seriously, I think we can say things like this. If Rob were to throw a wad of paper across the room, such that Shelia caught it out of the corner of her eye and lunged for it, maybe she *mistook* it for a 'tennis-ball.' That is, maybe she *targeted* 'that projectile' and applied her 'tennis-ball' *representation* to it. And when she caught it, she might even realize her mistake (and glare at Rob). Why not? [Haugeland in Clapin 2002, p. 141]

Dogs, too, according to Haugeland, have representations, and even if their contents are hard for us to specify, these are representational in the same sense that ours are. Indeed, on this view, the only way to explain Sheila's behaviour, whether accurately directed upon a tennis ball or mistakenly directed upon a paper wad is to say that she was guided by an internal representation of a tennis ball, which representation happens to match tennis balls and fails to match paper wads. Her mistake is one of comparison.

Now, there are mistakes of comparison: When I see a photograph of Brett and take it to be of Shane, I mistakenly apply a representation to that which it does not represent. I have compared it and mistakenly taken it to match. But not all mistakes are mistakes of comparison or of misapplication. When I add five and seven and get thirteen I do not misapply a representation of thirteen to twelve. I do not compare one with the other and decide that they match. And, closer to Sheila, when I swing wildly and miss because the ball was faster than I expected, I do not first represent the ball as here, and then swing; I see the ball and act. The relevant cognitive processes are tightly coupled to the perception-action cycle. The error in question is more like the failure of an arrow to hit a target than the misidentification of one target apple as another. So, while it might be that both error and success, and so all cognition, presupposes norms of correctness, it in no way follows that cognition presupposes representation whose success or failure to match that to which it is applied constitutes success or error.

What do we learn from this interchange? The unquestioned assumption that representation is the foundation of cognition leads to quandary: Representation *should* be decoupled from represented. Similar cognitive processes *should* be treated similarly by theory and metaphysics. Immediate sesnsorimotor engagement with the environment is a paradigm of non-representational activity, but yet seems continuous in important respects with a great deal of more reflective activity. So where is the line to be drawn?

7.   Historical Excursion

Berkeley famously attacks the idea of mental representation because of its entailment of a doctrine of double existence, a doctrine he finds unintelligible due to our inability to conceive of any *representeds* for our putative *representations* to *represent.* I don't want to worry about that argument, and I don't think that many today have much sympathy with the idea that the external world is really inconceivable.

But I do want to focus on one saner insight that Berkeley develops in the *Dialogues* en route to that conclusion: an insight he develops in the context of his discussion of the bust of Caesar. He *grants* that the bust *is* a representation, precisely because it presents to us what is absent, and more importantly, because it *mediates* our knowledge of that which it represents. To perceive Caesar, it would be best to have the Roman emperor himself in plain view. But alas, he is long dead, and the best we can do is to view his bust. That is a *way* of perceiving Caesar, but only a *mediate* way. Moreover, Berkeley notes, that mediation is made possible precisely by a great deal of background knowledge (importantly, for instance, the knowledge that it is in fact a bust of Caesar, and not, say, Cicero—a kind of knowledge he correctly, if for the wrong reasons, notes must be absent with regard to our inner states) and by a set of representational conventions—seeing the bust we do not take Caesar to be represented as a cold, white, legless monster, but as a flesh and blood Roman. Without the relevant knowledge and the conventions, representation of Caesar by the bust would be *impossible*; were we face to face with Julius, the representation would have been *unnecessary*. Representation is a mediate way of knowing the absent, in which mediation is achieved through the use of external props scaffolded by a network of relevant conventions and auxiliary knowledge. Berkeley's central and important insight then is that *perception* (which he might well

have understood in part in that older sense of the term, given his ecclesiastical calling) is *immediate communion* with its object, and so is the very *antithesis* of representation.

Hume is another early perceptive critic of the representational metaphor. He takes the doctrine of double existence to be a monstrosity, while pointing out both that the only things *immediately* present to the mind are its perceptions *and* that "'tis vain to ask whether body exists; this is something that must be taken for granted in all of our reasonings." The depth of Hume's position only emerges when we take seriously the task of rendering these claims—claims that to our ears, accustomed as they are to the trope of representation, are *prima facie* inconsistent—consistent.

While Hume is sometimes read as holding an imagistic model of thought, this reading is hard to sustain on a careful reading of the *Treatise*, where literal talk about the mechanisms of thought is almost always *causal* talk. Given Hume's disparagement of the idea of *double* existence and his insistence on *taking for granted* the existence of external objects, it is tempting to think that we can only understand him as arguing that there is no second *internal* existence of external objects as representations. In fact, things are more complex and more interesting.

Hume, at least as much as Berkeley, takes seriously the etymology of "perception" as originally denoting the act of receiving the Eucharist. Hume is not so much a foe of the *outer*, as naïve readings of the *Treatise* would have it, nor a foe of the *inner*, as a proto-behaviourist reading might have it, as he is a foe of the *very distinction* between inner and outer. In this regard he is very much more a post-modern figure before his time—a fellow traveler of Nietzsche and Heidgger—than he is a modernist to be lumped in intellectual history with Descartes, Malebranche, Locke and Berkeley. Hume, after all, is writing a treatise of *human nature*, and is very much concerned, as he puts it "to introduce the experimental method into the moral sciences." So transcendental ontology is far from his mind. For Hume, perception is not *in* the mind, nor is the object *outside of* the mind. Rather person, culture (this will become important later) and environment are continuous aspects of nature, and perception is but a rapport of the perceiver or thinker with its object—a communion. Perception then names the process or relation, not an inner state *re-presenting* an outer object.

It is Kant, in the first *Critique,* who really introduces the vocabulary of representation—*vorstellung*—into philosophy in the form we know it even if Descartes had already represented the "idea idea" as Rorty has called it. And it is Kant who transforms modern thought about thought by turning away from the *idea* as the unit of thought in favor of the *judgment*. These developments are, of course, closely intertwined.

Let us begin with the second of these contributions. Kant's insight that the fundamental epistemological unit is the judgment, the assertion that something is the case, as opposed to an idea, an unstructured something of which the mind is the subject, opened the way to seeing the contents of mind as having propositional structure, and so as having the logical structure of linguistic representations. This insight opens the way to seeing thought as involving deductive reasoning, and to seeing cognitive success as akin to the truth of utterances. Now utterances, at least the assertions with which Kant was primarily concerned as models of judgment, are plausibly at least often representational. As Wittgenstein was to put it much later, they re-present *states of affairs*. So Kant, by internalizing these linguistic representations with their logical structure, provided a clear analysis of just how representation might go, as well as a mechanism for representing *facts,* not *things,* and so of representational thought as *active thinking*, not the passive *having of ideas*.

This contribution becomes important in the present context for two reasons: First, in seeing thoughts as judgments, and so as akin to inner *sentences* Kant provides a compelling reply to the critique of the double-existence model of representation adumbrated by Berkeley and Hume. For on this account there is no second existence in the mind of the objects

of thought—trees, tables, numbers, golden mountains or unicorns—any more than there is a second existence of these things in sentences of German or English. Kant hence, by focusing on truth as cognitive success, as inferential relations as constitutive of content and sensitivity to norms as central to the cognitive, develops the first sentential model of thought. This proves to be an enduring model in the history of philosophy of mind in the West, and preserves the representational content of the idea idea down to the present. Secondly, by using *language*and the relations between sentences as the model for thoughts and cognitive processes, Kant directs our attention to what, as I will argue below, is the important germ of truth in the representational theory of mind: when we learn to think in sentences, we learn to think about representations, and hence, in a derivative sense, to represent.

Now back to the first contribution: Kant uses the term "representation" in three ways. In one sense, a representation is a mental episode, in the case of judgments, one that represents reality as being in a particular way; in the case of concepts or intuitions one that represents an abstract entity or a singular entity. In another, a representation is an object. Space is a representation, and so is a house or a boat. In short, all phenomena, as opposed to noumena, are representations in this second sense. In the third sense, representation is the *process* of representing a representation (in the second sense) via a representation (in the first). Representation is thus an activity of mind as well as a mental episode, as well as the object of representation.

It is important to see that Kant hence provides the foundation for the contemporary orthodoxy regarding the ubiquity of representation in cognition and regarding the linguistic structure of the posited mental representations. On the other hand, Kant's equivocation on just what a

representation is should give us pause. Do we need all three senses in order to get all of these cognitive benefits? While it is impossible to make sense of human cognition without taking the objects of thought sometimes to be representations (Kant's second sense, but in a restricted domain) and while that requires us to take seriously the activity of representation as one kind of thought, it requires us neither to see the *objects* of thought as re-presented, nor to see *all* of thought, and in particular, the *vehicle* of thought, as representational.

Wittgenstein in the *Tractatus* noticed that Kant's metaphor of representation included not only thoughts, but also the overt utterances that were their Kantian model. This insight is, of course, momentous for the subsequent course of Twentieth Century philosophy of language and philosophy of mind. The picture theory of language has, of course, been subject to trenchant criticism, as has the view that *all language* is representational; and indeed the view that representation is the foundation of linguistic meaning—and indeed the latter two critiques originate famously with Wittgenstein himself. Despite the fact that these theses bid fair to be the core of the *Tractatus* model of language, there is a significant baby to be strained from the bathwater, and indeed the baby destined to grow into the later Wittgenstein: it is not *thought*—something interior to the mind—that is the primary locus of representation, but *language*—a public activity, whose representational power is vouchsafed by rules of interpretation. The Kantian insight that representation is crucially dependant upon responsiveness to norms is retained; but the norms in question are norms of interpretation rather than exclusively norms of inference; while their essential publicity has not yet come to the fore, it is certainly in the offing.

In the *Investigations,* Wittgenstein explores the limits of the picture metaphor, an exploration sufficiently familiar that we can cut to the chase. The chase scene emerges at section 308:

308. How does the philosophical problem about mental processes and states and about behaviorism arise? The first step is the one that altogether escapes notice. We talk of processes and states and leave their nature undecided. Sometime perhaps we shall know more about them -- we think. But that is just what commits us to a particular way of looking at the matter. For we have a definite concept of what it means to learn to know a process better. (The decisive movement in the conjuring trick has been made, and it was the very one that we thought quite innocent.) -- And now the analogy which was to make us understand our thoughts falls to pieces. So we have to deny the yet uncomprehended process in the yet unexplored medium. And now it looks as if we had denied mental processes. And naturally we don't want to deny them.

We don't want to deny them. But precisely *what* do we not want to deny, and *why*? And what is the difference between accepting mental processes and accepting the "analogy which was to make us understand out thoughts?" Answering these questions takes us to the heart of the problem caused by the representation metaphor and allows us to glimpse its solution. It would be mad to deny that we think; that we have beliefs, and so forth. And it would be mad to deny that when we think and when we believe, we think and believe *about* things. Neither cognition nor its intentionality is up for elimination. But it is all too easy to move from thinking to thoughts, and then from talk about thoughts to talk about particulars that are thoughts. Having hypostasised these particulars, and having characterized them intentionally, we move right along, via the picture metaphor (or its linguistic cousin—there is no real difference here), to thinking of them as inner representations and hypostasising the processes in which they figure as computations over representations.

But now, if we have followed the argument of the *Philosophical Investigations* to this point, everything *does* fall to pieces: For we find ourselves trying to understand thought both as primitively

representational—as meaningful—and committed to understanding meaning as essentially rule-determined, and rules as primitively social. The "decisive move in the conjuring trick" was the reification—the hypostasy of specifically *inner* states and processes as the embodiments of thoughts and of thought. Accepting the mental, however, no more requires positing these inner entities standing for that about which we are thinking than accepting the fact that the arrow points left requires positing an entity *within* the arrow that *stands for* leftness.

The problem with the representational model of mind—independent of the particular model of representations—is that it takes thought about things to consist precisely in the deployment of inner tokens or processes whose purpose is to stand in for, rather than to enable rapport with, the objects of thought. The question that is left open—that of just how standing in for something enables thinking about it—is exactly the same question that representation was posited in order to answer in the first place: how can thought be about something. Pointing to an isomorphism is plainly inadequate and even irrelevant. The same questions arise immediately. And if we shift ground and talk about selection history or ability to guide behaviour it turns out that the representation relation has not been explained but rather abandoned. For nothing in these stories requires the doctrine of double existence.

5.   **Intending without representing**

As I noted above, much of my discontent with profligate talk about mental representation arose from reading Andy Clark's recent work even though Clark, at least in some moods, does not endorse my position. Consider the following remarks:

> The status of an inner state as a representation thus depends not so much on its detailed nature... as on the role that it plays within the system. It may be a static structure or a temporally extended process. It may be local or highly distributed. It may be very accurate or woefully inaccurate. What counts is that it is *supposed* to carry a certain type of information and that its role relative to other inner systems and relative to the production of behaviour is precisely to bear such information.

…[L]et us call a processing story representationalist if it describes whole systems of identifiable inner states (local or distributed) or processes (temporal sequences of such sates) as having the function of bearing specific types of information of external or bodily states of affairs. Representationalist theorizing thus falls towards the upper reaches of a continuum of possibilities whose nonrepresentationalist lower bounds include mere casual correlations and very simple cases of what might be termed "adaptive hookup." Adaptive hookup goes beyond mere casual correlation insofar as it requires that the inner states of the systems are supposed (by evolution, design, or learning) to coordinate its behaviors with specific environmental contingencies. But when the hookup is very simple (as in a sunflower, or a light-seeking robot), we gain little by treating the inner state as a representation. Representation talk gets its foothold, I suggest, when we confront inner states that, in addition, exhibit a systematic kind of coordination with a whole space of environmental contingencies. {Clark 1997, p 147]

> The most potent challenge to a representation-based understanding comes, we saw, from cases in which the web of casual influence grows so wide and complex that it becomes practically impossible to isolate any "privileged elements" on which to pin specific information-carrying adaptive roles. Such cases typically involve the continuous, reciprocal evolution of multiply tightly linked systems, whose cumulative ('emergent') effect is to promote some kind of useful behaviour or response. [Clark 1997, p166-167]
>
> The notion of internal representation thus gets a grip only when we can make relatively fine-grained assignments of inner vehicles to information-carrying adaptive roles. … At the very least, we can now see more clearly what it would take to undermine a representation-based approach: it would require a demonstration that, even in the representation-hungry cases, it remains practically impossible to isolate any system of fine-grained vehicles playing specific information-carrying adaptive roles. [Clark 1997, pp 168-p.169]

Clark certainly recognizes that a lot of cognition may involve simply what he calls "adaptive hookup," and that this in no way implicates representation. But he does argue that there is nonetheless a class of

"representation-hungry" cognitive processes that require us to posit a system of mental representations at least for complex mammals such as ourselves, and he asserts that the necessary conditions for a state's being a representation are that its function is to bear information about the environment and that it is a member of a family of "fine-grained vehicles playing specific information-carrying adaptive roles."

Now I will shortly deploy other remarks of Andy Clark against the position he defends here. But first I want to note some points with which I do not wish to quarrel: There are certain activities in which human beings engage for which representations are indispensable. And indeed for something to count as a representation requires that it carry information, or at least that its *function* is to do so and that it be a member of an articulated family of such information bearers.

But these conditions are plainly not sufficient for representation, as opposed to what Clark felicitously calls "adaptive hookup" and with which he contrasts representation. His examples of adaptive hookup include the processes that enable sunflowers to track the sun. Note that these states satisfy Clark's set of putatively necessary and sufficient conditions for representations. Without belabouring the point, or getting further into Chisholming away at sets of necessary and sufficient conditions for representation, we can get directly to the heart of the issue in this way: What is missing from Clark's account of representation is the very notion at issue—that of *standing in.* Carrying information and coordinating behaviour is indeed a crucial function of many of our cognitive states and processes, and indeed we can rank cognitive processes on some kind of scale of sophistication reflecting the fineness of grain required of the information-bearing states in question; the quantity of information required; the complexity of the processing involved, etc.

And it might make sense to distinguish in this way between *information-hungry* and *information-declining* processes in this way. But this does not get at the distinction between hookup and representation. For all the information in the world might not involve the *standing in* function.

In fact, Clark seems to be investigating not the distinction between hookup and representation, but rather grades of *intentionality*, and it is the distinction between *intending* in this sense and *representing* that is at issue here. The states to which Clark adverts, right from the merely hooking-up states of the sunflower to the complicated articulated states of a tennis-ball seeking border collie, and on into those of a human strolling through the art museum, *intend* their objects—the position of the sun; the ball that was just thrown, Monet and his role in world history. But none need *stand in for them.* None need re-present them. They need to coordinate increasingly sophisticated interaction, and do so in concert with increasingly complex families of states and processes. But that only gets us to higher degrees of intentionality. None of it constitutes representation, though once we hit the most sophisticated of these embodiments of intentionality, that of the art aficionado, for instance, it is clear that this intentionality presupposes a background of representation, a backgrounds to the conditions of the possibility of which we now turn.

6.   **Real representations: Institutions, Innovations and Language**

The truth is almost embarrassingly obvious, that it is human natural language that sets us apart from other species, because it gives us, and only us, for the first time, a genuinely open-ended compositional, manipulable, medium of representation suitable for any topic. [Dennett in Clapin 2002, p. 191]

Does this mean that we are approaching a nihilistic conclusion regarding representation? Not at all. We simply have been looking in the wrong place. Let us recall the lesson I suggested we learn from Wittgenstein: Representation is essentially a public, convention-governed phenomenon. For something to stand in for another—in any sense—requires that we have a set of rules for instituting and using the proxy. Just as I can stand in for the dean if I am appropriately deputized, and a cheque can stand in for a pile of money given appropriate banking institutions, a photograph can stand in for the grandchildren if we have a convention for reading visual information from a flat surface, and words can stand in for the picture if we speak the same language. (It might be thought that one of the central morals of the first half of *Philosophical Investigations* is precisely that, *contra* the *Tractatus*, the language is not representational. This would be a serious exegetical error—and one that would impute a serious philosophical error to Wittgenstein. In fact the point one might so misrepresent is that representation is one of the *many* functions of language, and presupposes, but is not presupposed by the more general practice of rule-following.)

The alternative approach is that we have seen taken by Clark, to take representation as a matter of bearing information in virtue of reliable covariation, and doing so in virtue of having that function (however that function is determined). So the neurons that bear information regarding the orientation of the rat's head, that provide that information to motor control systems, and that have precisely that function qualify in that sense as representations. Here is what is wrong with that approach: First, such systems incorporate no stable, context-independent recurrent states that would naturally be identified as tokens of representations. That's not a *big* problem, as one might simply begin a story about representation in dynamic systems. But it does begin to strain the applicability of the metaphor. Second, and more significantly, we can note that defining representations *this* generously gives representational status to the states of sunflowers, thermostats, and so forth. We should worry that we are about to lose the ability to draw distinctions that need to be drawn if we are to isolate the natural kinds of use to cognitive science.

Third, and now most importantly, we can point to the nature of that relevant distinction. We are after the distinction between *intentionality* and representation. It is certainly the case that the rat neurons to which Clark adverts, as well as the relevant structures of sunflowers are *about* that about which the carry information; that is to say, they are *intentional*. But to be intentional is not

sufficient to represent. Representation *is* important: it is what makes puts the *sapiens* in *Homo sapiens*. Intention is indeed a necessary condition of getting one ting to stand for another, but to get from intention to representation we need to go through convention, and thus through the enormous social intelligence for which we, alone among terrestrial species, are specialized. Clark simply sets the bar too low, and misses the importance of this distinction.

Representation is *symbolic,* and the use of symbols is an innovation requiring norm-imposing and norm-enforcing institutions that make it possible for rules to be followed and hence for proxies to be employed. The pre-eminent and indeed, for humans, omnipresent, system of such representation is, of course, natural language—an innovation so critical to our species that we have become specialized for it, breeding out those who cannot cope easily with it and selecting for those with linguistic facility. The ubiquity of linguistic representations and the ambient bath of norm-governed practices enabling and enabled by it is so great that like water for the fish it is invisible to us. We use linguistic representations when we speak, when we read, when we listen, and when we think. And when we use language, just as when we use pictures or cheques, we can allow symbols not only to intend, but to *stand in* for that which is absent (though, to be sure, that is not *all* we do with language). Here is Andy Clark again:

What does public language do? There is a common, easy answer, which though not incorrect, is subtly misleading. The easy answer is that language helps us to communicate ideas. It lets other human beings profit from it from what we know, and it enables us to profit from what they know. This is surely true, and it locates one, major wellspring of our rather unique kind of cognitive success. However, the emphasis on language as a medium of communication tends to blind us to a subtler but equally potent role: the role of language as a tool that alters the nature of the computational tasks involved in various kinds of problem solving.

…

Public language is in many ways the ultimate artifact. Not only does it confer on us added powers of communication; it also enables us to reshape a variety of difficult but important tasks into formats better suited to the basic computational capacities of the human brain. Just as scissors enable us to exploit

our basic manipulative capacities to fulfill new ends, language enables us to exploit our basic cognitive capacities of pattern recognition and transformation in ways that reach out to a new behavioral and intellectual horizons. … Finally, the sheer intimacy of the relations between human thought and the tools of public language bequeaths an interesting puzzle. For in this case, especially, it is a delicate matter to determine where the user ends and the tool begins! [Clark 1997, p 194]

Public language is indeed the ultimate artifact. And indeed its great importance is *not* is power to convey that which we could already think. There is, I am afraid, not much of *that* worth conveying—though there is some, to be sure. The great value of the ultimate artifact—as Dennett puts it, that which sets us off from all other species—is that it makes possible representation in the full-blooded and fully versatile sense. And this makes possible the human thought we know, love, and really do wish to convey. We do not simply coordinate our behaviour with our fellows or with our non-human environment; we acquire the ability to manipulate in thought, that is, to reason about, to imagine, to plan regarding, things that are not present to us; to categorize and re-categorize using labels, that is, to conceptualize, both the concrete and the abstract in our environment. This is the true dawn of representation. And here we see the intuition that there is something very special about representation vindicated. The ability to stand in for is more than the ability to mediate purposive behavior. Ignoring that function in favour of mere intentionality is perilous to the philosophy of mind.

But what does this mean regarding the idea of *mental representation?* Here is what I think it means: There is none, really. We do represent, and we use our minds to do it, but only in a derivative sense. The representational burden is carried by the tokens of our public language. They stand in for their denotata when we represent. What our mental processes are good for is intending. But with the development of the institution of language, we acquire the ability to intend any of a vast

number of articulate representations. Mental intentionality plus linguistic representation equals human thought.

This picture is buttressed by important evidence from developmental psycholinguistics, particularly when those data are viewed from a broadly Vygotskyan perspective. The principal data I have in mind come from research into the acquisition of theory of mind, and in particular regarding the role of pretence in that process. But before I turn briefly to those data, let me note that Clark's intuitions converge. Reflecting on recent work by Berk and his colleagues, he writes:

> The Vygotskian image is supported by more recent bodies of development research. Berk and Gavin (1984) observed and recorded the ongoing speech of a group of children between the ages of 5 and 10 years. They found that most of the children's private speech (speech not addresses to some other listener) seems keyed to the direction and control of the child's own actions, and that the incidence of such speech increased when the child was alone and trying to perform some difficult task. In subsequent studies (Bivens and Berk 1990; Berk 1994) it was found that the children who made the greatest numbers of self-directed comments were the ones who subsequently mastered the tasks best. Berk concluded, from these and other studies, that self-directed speech (be it vocal or silent inner rehearsal) is a crucial cognitive tool that allows us to highlight the most puzzling features of new situations and to better direct and control our own problem-solving actions.
> [Clark 1997, p 195]

Data from our own laboratories, those of Thomasello and his colleagues concerning social development and the findings of the de Villiers and Senghas and Pyers corroborates this intuition in a dramatic way: The de Villiers demonstrated dramatically that the development of competence in reasoning about mentalistic states immediately follows and requires mastery of the syntax and semantics of the sentential complement constructions and the verbs that govern them. Without this linguistic competence, it is impossible for children to reason about psychological states. Senghas and Pyers have shown that this inability to pass theory of mind tasks without mastering the language necessary for representing

those phenomena. Our own research has shown that pretence plays a special role in coming to be able to reason about mental states: children master verbs of pretence and the ability to reason about pretence prior to attaining corresponding mastery of the language of the mental and of the ability to reason about the mental.

This pattern—the learning the relevant language preceding the ability to perform the reasoning; the precedence of mastery of reasoning regarding joint pretence over that regarding inner episodes—confirm the Vygotskyan intuition that representation is initially a public activity and that as the public representational medium of language is learned, it can scaffold autonomous thought and reasoning. The most plausible mechanism of this scaffolding is that our ability to intend is turned upon this representational medium, allowing us to make autonomous use of language. The representational weight is born by language; the ability to make use of that power in thought requires our complex cognitive ability to intend—to engage cognitively with our environment. When that environment includes symbols, the magic of human thought occurs. Clark puts the point this way:

> Experience with external tags and labels thus enables the brain itself, by representing these tags and labels, to solve problems whose level of complexity and abstraction would otherwise leave us baffles-an intuitive result whose widespread applicability to human reason is increasingly evident. Learning a set of tags and labels (which we all do when we learn a language) is, we may thus speculate, rather closely akin to acquiring a new perceptual modality. … And of course the whole process is deeply iterative—we coin new words and labels to concretize regularities that we could only originally conceptualize thanks to a backdrop of other words and labels. [Clark in Clapin 2002, p. 42]

Note that while Clark accepts the endpoint—the claim that it is language that enables the full range of human representation, he continues to elide the distinction between intention and representation, requiring the brain not merely to *intend*, but to *represent* the tags and labels, requiring of language a new version of the doctrine of double existence. It is easy to see why we are tempted in that direction: if language provides our metaphor for understanding thought, and if language is the example *par excellence* of a general representation scheme, it is natural to import that feature of the metaphor to our model of

thought. But precisely the point of our double adaptation for and to language is to save our brains the trouble of representation, by leaving that task to the enriched symbolic environment we construct and inhabit.

One final point needs to be made before we leave this topic. While I have been urging that the representation enters cognition through the medium of language, it does not follow from this that all language is representational, or even that the fundamental function of language is to represent. That was the error of the picture theory. *Some* of the uses to which language is put are representational, and it is these that scaffold the derived representational character of *some* thought. But a great deal of what we do with language, as Wittgenstein and Vygotsky in different, but related ways, have each emphasized, is non-representational, and the very ability of language to represent itself is the outcome of a range of conventions that, on pain of regress, must be conceived as *prior* to representation.

7. The abstract, the distant and the nonexistent, but also the general, the motley and the meta

Discussing with approval Dennett's view of the role of language in representation in Hugh Clapin' excellent conference proceedings on mental representation—Clark edges closer to this position:

> Florid representing occurs, recall, when there is a *knowing* use of *representations*, where representations are (at least) some kind of maipulanda: objects that bear contents and that can be somehow shuffled, reorganized, and recombined in ways sensitive to, and exploitative of, those contents). … The objects (the manipulanda) involved in florid representing bear the contents they do only in virtue of a bedrock of skills and capacities, rooted in multiple non-propositional mind-tools. But florid representing depends on making those skill-based contents into objects suitable for the exercise of other (non-propositional) skills— skills of combining, shuffling, and so on. And it is this 'objectification' of certain aspects of content that supports the highly versatile and open-ended range of thought characteristics of (and perhaps uniquely characteristic of) human understanding. Finally, … it is our experiences with public symbols that are said to teach us to make more manipulable objects of our thoughts and ideas [Clark in Clapin 2002, pp 81-82]

Here we see the recognition that (1) representations are *manipulanda*—the objects of cognitive operations, and not their constituents; (2) that their content depend upon non-propositional skills; and (3) that public language plays a crucial role both in generating human representational power and in generating the possibility of thought about thought. This is all to the good. All that is

missing is the recognition of the deep connection between (2) and (3)—that the basis on which representational content is constructed comprises *both* the non-propositional intentional capacities and skills of the organism *and* the representational capacity of public language, joined in the ability to think about language, and derivatively about that which it represents.

Does this claim about derivative mental representation threaten to collapse the distinction between intention and representation? One might think so, since on this view when I think, for instance, that if I were now in the northern hemisphere I would be in summer, I think about the northern hemisphere and do so via a representation, and hence the relevant thought is representational. And since there is nothing of ontological significance to distinguish between the nature of this thought and the states of my perceptual or motor control systems that I want to point to as paradigms of the sub-representational intentional, the distinction collapses. But this is not the end of the story. The representation in this case is, crucially, not the *vehicle of thought,* not a cognitive state, but rather the *immediate object of thought,* a linguistic item. Cognition *intends* that object, which in turn *represents* the abstract fact in question. This derivative representational character also allows us to understand how the continuous processes of cognition can have discrete semantic content. Their continuity is resolved by the discrete objects they intend. The kernel of truth in the language of thought hypothesis is the intuition that representation must have determinate, and indeed, compositional, content, and that only language can provide that. It does not follow, however, that thought is *in* language, only that it is *of* language.

It is a commonplace in contemporary foundations of cognitive science that language is a necessary condition of the representation of the non-existent, the abstract and the distant. We can now reformulate this point: representation is a necessary condition of the intention of the non-existent, the abstract and the distant. Without the capacity to represent, one simply can't think about these things at all. But once one can represent, the sky isn't even the limit.

To this trio, we can add another trio not often noticed: the general, the motley and the meta. Dennett notes the point about generality:

> Animals are not only capable of believing general propositions *in sensu composito.* That is, consider that case of learning that all orange mushrooms are toxic. What the dog can learn, perhaps learning from training from its mother, is

the disposition, whenever you see an orange mushroom, to shun it. That is, the disposition to acquire a particular 'this is toxic' belief, whenever encountering a particular orange mushroom. [Dennett in Clapin 2002, p. 50]

An infralingual creature can surely reliably intend each member of some class on each occasion, and even be said counterfactually to do so. In one sense this might count as having a general attitude. But, as Dennett points out, this is a kind of amphiboly: it is instead generally true of such an animal that it intends each member in the same way. Generality emerges with quantification, and quantification is a linguistic device.

We can believe things of motley collections, for instance that each item either on my desk or once owned by GE Moore is beautiful. This power derives specifically from the representational power of language, and outstrips any power of mere intending. It is central to the full generality of thought. Finally, we can think about thought, both in virtue of intending our thoughts and in virtue of representing them. Thinking about thought requires the kind of iteration of intentionality that is made possible only by representation, precisely because it is only representation that gives us a description of thought as intentional, and so makes it possible for thoughts *qua thoughts* to be objects of thought.

To be a human thinker is precisely to be able to think in full generality, about just anything, and to reflect on that capacity. The gap between regular adaptive coupling with the local environment and this capacity is precisely the capacity between intentionality and representation that is made possible by language.

8.   Not just a notational variant!

It may be tempting at this point to reply to all of this that I have only rehashed the difference between linguistic and non-linguistic representation—that my term "intending" is nothing but a notational variant of "non-linguistic representation." This would be a grave error. I am neither after the point that language makes mental representation more complex and powerful than non-linguistic mental representation; nor am I after the position that infralingual animals have no

intentional states. These are the two extreme positions between which I seek to navigate. The former position commits the error of supposing that all (or at least a lot) of infralingual cognition is representational at all. Again, if we pay attention to what is involved in that commitment, that is the commitment to a doctrine of double existence, and that is both unwarranted and non-explanatory. It distracts us from the task of explaining organism-environment coordination and leads us to look for sets of inner tokens that stand in for the distal, together with operations on them instead of searching for processes that guide interaction with what is already present externally. This involves a confusion of intention with representation. The second position begins in the same confusion but takes it in the opposite direction. Running together intention and representation, the eliminativist of this stripe (say Davidson in his denial of thought to animals) notes that since infralinguals lack representation, they must also lack all intentionality. Each error can be avoided simply by drawing the distinction in question.

The real point, then, is that in the search for the joints at which to carve cognition in the butchery of cognitive science, there is positive gain to seeing the emergence of representation as the social achievement of *Homo sapiens* in the evolution of and for language, and to understand more basic cognition, including the human cognition that subserves the representational, as intentional, but *not* as representational. Dennett puts a very similar point this way:

> The upshot is that believing is pervasive and fundamental. But human-style conscious awareness requires an extra layer of judgment rooted in a culturally inculcated capacity to spin a privileged report or narrative: "the story you or I will tell if asked (to put a complicated matter crudely)." (Dennett 1995: 348).

This remark reflects an additional related insight: It is with representation, and hence with language, and hence with complex social organization, that narrative self-consciousnes and introspective awareness—that is, as Kant would put it, experience—becomes possible. This analysis delivers the conclusion that the first person singular of the *cogito* presupposes the first person plural of community membership.

9.   **The real problem: getting the outside in**

What is really wrong with the whole representation metaphor as a way of understanding cognition, *per se?* Just this: We want to understand how we can perceive, act on, remember our environment—the stuff *outside* of the nervous system. We decide that the only way to explain that is to get it all *inside* the nervous system, in the form of a set of symbols. We can then act on those, and derivatively act on *it*. It is as though we are inside our house and want to wash the car. How do we do it? Make a model of the car, wash it with pretend water, and be done. It doesn't work. We need some connection between the model and the car, between our actions and the actions to be performed out there. And that's just where we started.

We still need an explanation of how our symbols actually *connect* to the environment, and how our *inner* operations actually *act* on the environment. And that's just where we started. It makes so much more sense, evolutionarily, cognitively and philosophically, to leave the external world outside, and to work out how natural organisms coordinate their inner lives with their external environment directly, saving the story about re-presentations for the part where representations become part of that larger environment. That final ornamentation of thought we take so much for granted is indeed a momentous occurrence, but an ornamentation that can be added only once a very large cake has been iced.