

## Dissolving the Grounding Problem: How the Pen is Mightier than the Sword

Nancy Salay (salay@queensu.ca)

Department of Philosophy and School of Computing, Queen's University,  
Kingston, Ontario, Canada, K7L 3N6

### Abstract

The computational metaphor for mind is still the central guiding idea in cognitive science despite many insightful and well-founded rejections of it. There is good reason for its staying power: when we are at our cognitive best, we reason *about* our world *with* our concepts. But the challengers are right, I argue, in insisting that no reductive account of that capacity is forthcoming. Here I describe an externalist account that grounds representations in organism-level engagement with its environment, not in its neural activity.

**Keywords:** computation; representation; externalism; intentionality; language; embodied; dynamic systems

### Introduction

The computational metaphor for mind is still the central guiding idea in cognitive science despite many insightful and well-founded rejections of it (Chemero, 2009; Dreyfus, 2007; Keijzer, 1998; VanGelder, 1995; Clark & Toribio, 1994; Freeman, 1990). Indeed, it is the central theme of this year's conference. There is good reason for its staying power: when we are at our cognitive best, we reason *about* our world *with* our concepts, unemotionally, amodally, and according to formal principles. And yet the metaphor leads to a deep puzzle for which there is as yet no plausible solution: how does this computational/representational (CR) capacity emerge out of the neural-body substrate that makes us up? The response to this theoretical gulf has been, more or less, two-fold: on the one hand, those focusing on embodiment, environment, and on the dynamic interplay between complex systems, have become suspicious of the starting intuition, that cognition is computational at all; and, on the other hand, those unwilling to let those intuitions go — and this is still the majority in cognitive science today — have tried to bridge the explanatory gulf by pushing the metaphor into service at the level of implementation, the brain. This move, so subtle that it tends to go unnoticed, is a mistake, a dangerous one, because it obscures the pressing area that needs attention, namely, the ongoing problem of intentionality, and polarises the field in a way that is unhelpful.

Here I will argue that although the computational metaphor plays a critical theoretical role at the explanatory level of mind, an internalist reductive account of its implementation just isn't available; consequently, we will need to begin developing an externalist account of this capacity instead. More specifically, the paper is organised as follows: in Section I, I clarify the scope of the argument; in Section II, I describe the slide down the explanatory ladder; in Section III, I argue that the only justification for the move, the promise of an internalist reduction of the relevant concepts — representation and computation — is not forthcoming; and,

finally, in Section IV, I provide a sketch of an alternative, externalist account.

### Section I: Setting the Terrain

Before heading into the discussion proper, an important terminological matter needs settling. While the cognitive/non-cognitive divide might be fuzzy, there is much cognitive activity, perhaps a majority of it, that does not necessitate a CR theoretical account. The constraint “when we are at our cognitive best” was meant to carve off that small subset of activity that does. For example, a full account of the cognitive mechanisms that make it possible for me to imaginatively plan my trip for next week will *require* CR concepts: it will have to explain how I represent to myself possible future scenarios, how, on this imaginative basis, I make a plan to take a certain set of actions, and finally, how, when the event finally occurs, I act on these plans. On the other hand, an explanation of how I manage to avoid an obstacle in my path will not require an appeal to such concepts: here the explanation can stay at the level of sensory receptors, motor responses, inhibitory activity, and the cascading effects of myriad sub-microscopic events.

Given that the cognitive terrain is so broad and that many of the terms already in use carry so much baggage, I will stipulate a term here, “R-activity,” in order to conveniently and unambiguously refer to just that subset of cognitive activity that *necessitates* CR descriptions in the sense described above. To repeat, I am not here making the case *that* the computational metaphor for cognition is a theoretically useful one; rather, I am beginning with the assumption that it is, but only for a small part of the larger cognitive landscape. It may be that we won't agree on the extension of R-activity, but this should not stop us from trying to become clearer about how brains could possibly implement it. As I note above, it is certainly the case that not all, perhaps not even most, of cognition constitutes R-activity. Humans ‘reason’ illogically, particularly where probabilities are concerned and even when they are prompted explicitly to make rational inferences (Tversky & Kahneman, 1974). Much of our waking life is not very self-conscious at all, that is, we carry on without much inner awareness of our own occurrent experiences and with very little ‘top-down’ control over what thoughts flit in and out of what little awareness there is. Finally, there are many features of self-consciousness itself that are not well-described within a CR framework: we have feelings, moods, emotions, interoceptions, all of which seem best explained in terms of complex systems concepts.

The starting assumption here, then, is the claim that any theoretical account of R-activity will necessarily involve CR concepts. In other words, if some phenomenon can be

comprehensively explained without appeal to such concepts, it isn't R-activity. The internalist reductive hope of the mainstream approach is that thin versions of CR concepts, used to describe activity at the neural level, can be systematically fleshed out to become the thick concepts requisite for an account of R-activity. I turn now to a discussion of how this thin-thick transition is supposed to work and why it won't.

## Section II: Down the Slippery Slope

There is a subtle, one might even say, sly, move down the explanatory hierarchy from computational level discourse to implementational detail. This sleight of hand is effected by a largely unnoticed equivocation on the terms "representation" and "computation." In a paper that exemplifies this shift, Edelman proclaims that both perception and action are computational activities *at the level of implementation*: "The cat's safe landing on the table indicates that *its brain* has successfully carried out a series of arithmetic operations on quantities *representing* ... the cat's mass, the dimensions of the cat and of the table, and their relative positions." (italics added) (Edelman, 2008) Critical here is the assertion that this computational description of brain activity is not merely one among many possible descriptions of it, so that we might consistently view the processes as non-computational within a different theoretical framework; rather, Edelman is here making the strong claim that this description is *theoretically explanatory*, that it is, in our terms, R-activity: "these two examples, along with many others in perception, memory, language, thinking, and consciousness, all point to the same conclusion: no matter how else cognition can be described, computation is what it actually is." (Edelman, 2008)

Now, as stated above, while we agree that CR concepts are required in order to explain, for example, how agents reason about situations that are not sensorily present, this is not the case for explaining action and perception more generally; some excellent, non-representational, accounts of precisely this are being developed (Noë, 2010; O'Regan & Noë, 2001). In these cases, while CR descriptions *can* be given, they are *merely* descriptions and not, as Edelman puts it, "what the activity actually is." Indeed, as I will now argue, there is no theoretically justifiable way to use those terms to describe the brain activity that underwrites action and perception more generally.

## Section III: Reductions Won't Work

First let's take a closer look at how the reduction is supposed to work. Neural activity 'represents'<sup>1</sup> some feature of the environment for an agent just in case 1) the neural activity nomicallly co-varies with the presence of the feature in question; and, 2) this co-variance relation *explains* the

functional role of those neurons within the organism.<sup>2</sup> For example, the activity of a set of neurons in the vomeronasal organ (VNO) of many mammals nomicallly covaries with the presence of certain pheromones in its environment. Being sensitive to the presence of these pheromones has survival value for the species because the presence of those pheromones is also nomicallly co-related with the presence of potential mating partners. In other words, it is *as a result of* the information relation between the presence of pheromones and the presence of potential mates that the VNO neural sensitivity developed at all. Thus, the co-variance relation between VNO neuron activity and the presence or absence of potential mates explains the role of those neurons for the organism: they function to detect or represent (to it) potential mates. Or so the story goes. But there are serious problems with such accounts.

In order for an information relation to be a factor in a functional role account of a mechanism, such as VNO pheromone detection, *information*, in this case, *the presence of a potential mate*, must be an integral part of the explanation. If the account can be given without appeal to this information at all, then information isn't playing a theoretically relevant role in the mechanism. The question then, with respect to information-theoretic accounts of neural activity, is where, at the neural level, information is playing this kind of integral role.

Neurons respond electro-chemically to certain stimuli, varying according to neuronal type and to the idiosyncratic experiential history of the larger organism. Whatever the distal stimuli are, the proximal stimulus of a particular response is an electro-chemical charge that either triggers or inhibits a neuron from releasing its own electro-chemical charge. This activity, in turn, either triggers or inhibits impulses in connected neurons in a stimulus-response chain reaction. Information abounds here, *as it does everywhere*, but in order to exploit a particular informational relation, as a human might do when she notices that the mail is delivered at 1:00 each day, the exploiter must be capable of responding to this higher order information relation. Individual neurons don't seem to have this 2<sup>nd</sup>-order capacity: they are embedded within a cause-effect web in which sub-micro particles are the primary players. Individual neural activity cannot be representational *for other neurons*, then, because, at the level of neurons, there is just a stimulus-response flow of electro-chemicals. There is no exploitation of relations for information here.

If we suppose that there must be a special class of neurons, say fore-neurons, that, *somehow*, are capable of exploiting these information relations, we would indeed have posited something capable of using the information embedded in neural activity, but only at a greater cost: the infinite homuncular regress problem. In order to explain the

---

<sup>1</sup> This 'thin' notion of representation is itself deeply problematic: the compressed dyadic relations of information theory are simply not representations in the relevant sense since there there is no 'interpretant' of the representation. Since this argument has been

thoroughly developed elsewhere (Salay, 2016; Salay, 2014; Ramsey, 2003), I omit it here.

<sup>2</sup> Following Dretske's (1988) information theoretic account of representation, arguably the most widely accepted and used such account.

intentional capacity of those fore-neurons, how they are able to make a connection between one thing, say a pattern of neural activity, and something else, say the presence of a potential mate, we'd have to appeal to some still smaller fore-particle inside of it. And so on. Appeal to an inner intentional agent to explain global intentional capacities never explains anything; it merely pushes the problem one level down.

A third possibility is that it is the organism, as a whole, that exploits the information relations of its neural activity. This seems to be the idea implicit behind the received view of sense perception. Individual neural sensitivities to different sorts of stimuli, connected across networks of different layers of neuronal structures, responsive to different *kinds* of stimuli, some internal and some external, together make up organism level perception. On this view, the fact that low level stimulus responses reliably correspond to important (for the organism) informational states explains why these sensitivities evolved in the first place. Using our earlier example, the presence of certain pheromones is a reliable indicator of the presence of potential mates and that in turn is a valuable piece of information *for the organism*. The organism, in being tuned to these lower level responses, in moving towards areas in which pheromone levels are higher and away from those in which they are lower, seems to be doing precisely what we are debating here, that is, *using* the information that a potential mate is present.

Unfortunately, the sense in which this sort of behaviour can be described as 'information use' is metaphorical at best. The deep problem here lies in the misguided idea that an organism could ever *use* information provided by its neurons: organisms don't *use* their own inner states at all; they are *constituted* by them. Indeed, at no point is an organism ever aware of the relationship between its own neural responses and the information about its environment that these responses carry, in the way that *we* have just become aware of it in telling the evolutionary story above. Craver and Bechtel develop this point in their excellent paper (2007). There they explain that the 'mystery' of top-down causation is really just a confusion of constituency relations with causal ones. Interlevel relations, these are between entity wholes and their mechanistic components, cannot be causal because the players in a causal relation must be, at a minimum, spatiotemporally distinct from one another: "If one of the parts bears a mark, that mark is always already born by the whole (by virtue of being born by its parts). The marks do not need to be transmitted upward or downward to have their 'effects;' their effects are inherited constitutively, not causally." (p. 552)

The deep problem, then, with accounts that take the sensory processes at the neural level to be *causally* responsible for the perceptual processes at the organism level, is just this mistake: treating interlevel constitutive relations as though they were intralevel causal ones.

The central feature of cognition we want our best cognitive theory to explain is, arguably, our capacity for representation,

---

<sup>3</sup> He describes it as a capacity to solve "representation-hungry problems." [cite]

for exploiting information relations in the ways in which anyone who is reading these words is doing. But, finally, in our zeal to give a naturalistic account, we risk diluting it to such a degree that it no longer serves as an account of the very thing we are so keen to explain, namely, *bona fide* representation use. Causation abounds in the universe, but representation, intentionality, is rare. The skin on my body is sensitive to different forms of light: when strong UV rays hit my skin, the pigment cells react. This response is one way of protecting the organism that is me from certain kinds of damage. Because it is quite successful in this, the adaptation has been passed on genetically across many generations of organisms and across species. We are not tempted to call these skin cells *detectors* or as functioning to *represent* anything at all, though from the evolutionary vantage point, they are acting no differently from VNO cells. In other words, it is not the activity of a given cell that makes us think it must play a part in a representational process; it is what we think is the end of such processes that compels us to give an information-theoretic account of some cell activity and a mechanical account of others. But this kind of teleological explanation is as unwarranted here as it is in scientific explanations more generally.

If brains alone can't be the ground of R-activity, then it looks as though we need to cast our theoretical net more broadly, as Andy Clark (2015, 2008, 2006a, 2006b) and others (Wheeler, 2007, 2004; Meneary, 2007; Rowlands, 2006) have been advocating for some time now. In the next section I sketch the groundwork for such an account.

## Section IV: The Language Tool

Clark, in recognising both the inherently representational nature of what I am calling R-activity<sup>3</sup> and the insights of dynamic systems, embodied challenges to mainstream cognitivism, has been developing a middle path (Clark, 2006, 2006, 2002, 1997) between the two extremes. While I am sympathetic with his motivation, I am not optimistic that his approach, to replace an internalist account of objectivist representations with an internalist account of action-oriented ones (Clark 1997)<sup>4</sup>, will do for the reasons rehearsed in Section III. Dreyfus' remarks (2007), though they don't address the problem of how we manage to achieve R-activity at all, ring true here: "Any attempt to solve the frame problem by giving any role to any sort of representational states, even online ones, has so far proved to be a dead end. It looks like nonrepresentational neural activity can't be understood to be the "extreme case." Rather, such activity must be, as Heidegger, Merleau-Ponty and Freeman contend, our basic way of responding directly to relevance in the everyday world, so that the frame problem does not arise." (p. 1158) The suggestion on offer here is that language itself is the locus of our representational capacities.

Dreyfus, in laying out the basic requirements of a Heideggerian AI, combines phenomenological insights about

<sup>4</sup> But see Clark 2015 for what might be the beginning of a retraction of this position.

significance for agents with Walter Freeman's (2000) anti-representational interpretation of neural activity. Critical here is the concept of significance: "The cell assemblies that are formed by the rabbit's response to what is significant for it are in effect tuned to select the significant sensory input from the background noise. For example, those cells involved in a previous narrow escape from a fox would be wired together in a cell assembly. Then, in an environment previously experienced as dangerous, those cell assemblies sensitive to the smell of foxes would be primed to respond." (Dreyfus 2007, p. 1152)

Clark has, rightly, complained that *significance* here sounds like a watered-down version of the cognitivist's *representation*: "it has never been clear, to me at least, why the various attractor states of the olfactory bulb (or whatever) do not count as internal representations of something like "smells-in-the-context-of-typical-actions." (Clark, 2002, p. 386) One problem here is that in being so steadfastly anti-representational, Dreyfus' account does not speak to R-activity at all and so one is tempted to search for something in his explanation that could. The second problem is in the implicit internal/external dichotomy that some of his descriptions, e.g. "...what is significant for it ..." (ibid) set up. These lend themselves to representational readings.

To avoid these problems, I will constrain talk of significance to what I will call "significance responses" (SRs): any neural responses to stimuli that are also constitutive of larger-scale organism-level responses — secretion of saliva, increases in heart rate, locomotion, perception, and so on. Many different types of stimuli might trigger similar SRs in a given organism (depending of course one how we draw our boundaries). An anxious person, for example, can become anxious as a result of a wide variety of circumstances.<sup>5</sup> The central point to note here is that SRs are ubiquitous.

Some SRs are innate, that is, genetically-encoded, while others are learned responses to local environmental conditions, and there is a continuum of interaction between these 'hard-wired' and learned behaviours. The degree to which an organism has a greater balance of learned vs innate SRs is expressed by the concept of *adaptability*: the more adaptable an organism is, the more its SRs are learned, that is, formed and influenced by local environmental conditions; the less adaptable an organism is, the more its SRs are innate, that is, are reflexive behaviours. Humans, for example, are highly adaptable since they have very few innate SRs — only some reflexive behaviour such as sucking and finger grabbing that is prevalent especially in newborns. Rabbits, on the other hand, exhibit both innate and learned SRs: whenever a potential predator is sensed, a flight response is initiated; however, what local conditions constitute a potential predator, can be learned. A rabbit might develop an SR (e.g. a flight response) to fox smells because it had a threatening fox experience in the past. The fox smell then

becomes a trigger of the response, a significance response trigger (SRT). As noted above, there might be multiple SRTs for a given SR — fox smells, fox vocalisations, fox silhouettes in addition to eagle smells, coyote barks, and so on. But, so far, there is no representation here at all, neither internal nor external, only a high-level account of how and why organisms are pushed and pulled in their environment.

Sometimes one bundle of features — bundled because they belong to some type of thing and so naturally always appear together — is co-related with another bundle of features — perhaps because there is a causal relationship between the two so that when one is present, the other follows soon after. To use a well-worn example, fire is almost always accompanied by smoke and smoke is often an early indicator that fire is present. Presumably, if an (adaptable) organism develops an SR to the presence of fires, so that fire situations become an SRT for it, the organism will also exhibit the same SR, though perhaps to a lesser degree, when only smoke is present. In other words, smoke, being one of the elements of the feature bundles that (often) accompanies fire situations, will serve to trigger, to some degree, the entire SR itself.

From the perspective of an organism being pushed or pulled to respond in some way to its environment, this observation that smoke is only sometimes present when fires are, that they are separate *things*, is not available; something either triggers the response, to a greater or lesser degree, or it does not. But from the perspective of theory development, we can see that there is a difference between fire as an SRT and smoke as an SRT: it is *because* fires pose a threat to an organism that it develops an SR to it in the first place — this explains why fire is an SRT at all; but, it is *because* smoke (typically) accompanies *fires* that an organism develops a response to smoke. Smoke here is a response trigger, but it is a response trigger for fire, which itself is the *raison d'être* of the original SR. In other words, it is an SRT once removed, a significance response trigger trigger (SRTT), we might say. Smoke, as Grice (1989) famously observed, is a *natural* sign of fire. In terms of the concepts I am developing here, we can say that smoke is a *natural* SRTT: it triggers, at a lesser intensity, significance responses that have developed to the presence of fires.

This relationship between SRTTs and SRTs can be exploited. If one wanted to elicit a particular SR, say a flight response normally triggered by the presence of fire, one could use smoke, an SRTT, which would result in a response *as though* to fire. We need not cash out this *desire* to invoke a flight response in others in propositional terms, of course. Dreyfus à la Merleau-Ponty is right to insist that, "as I cope with a specific task in a specific situation, other situations that have in the past been relevant are right now present on the horizon of my experience as potentially (not merely possibly) relevant to my current situation." (Dreyfus, 2007, p. 1158) In other words, there is a pull to do whatever can be done in the current situation, to use whatever is 'ready-to-hand.' But

---

<sup>5</sup> Developing a taxonomy of SR types would be theoretically useful, but I will leave details of that sort out of this high-level sketch.

even though there is no attendant inner understanding, no conscious “I am using this to do this,” this SRTT exploitation is nevertheless an example of representation use: unlike the relation between VNO neurons and potential mates, here the smoke is being *used* to trigger the response that fires generally trigger, and in this use, it is *standing in for* the fire itself.

SRTTs are sometimes behavioural elements themselves of the SR to be elicited. Some animal alarm calls seem to work like this. “Flagging,” when an animal lifts its tail to display the white fur underneath, serves to trigger the flight response in nearby animals, presumably in virtue of the fact that the tail is lifted and the white fur is exposed during the flight response as well, i.e. as the troupe of animals is fleeing. Of course “flagging” and other alarm calls are typically themselves innate SRs (to some potential danger situation) that function here as SRTTs in the context of a larger group of organisms. These are legitimate examples of information-exploitation because the SRTT elicits behaviour in the larger group, e.g. a flight response, that *would* be triggered by some SRT, some actual danger, were it sensorily available to members of that group. In this way, the SRTT is acting as a symbol or sign of the original SRT.

The more sensorily distant an SRTT is from an SRT, the more explicitly it must be learned. Smoke is a feature that naturally occurs when fires are present; the white fur underneath a raised tail is a natural feature of a ‘fleeing’ situation. This kind of symbol use can be ‘hard-wired’ precisely because it trades on physiological responses to the physical features of the SRT itself. But organisms that are adaptable, that are capable of learning new SRs to new SRTs, can learn increasingly abstract SRTTs. Bonobos, for example, have a wide repertoire of gestures and vocalisations, some of which seem to be grounded in the sort of natural SRT to SRTT relation described above, such as alarm calling, and others that seem to be locally learned, ontogenetic ritualisations that function as stand-in triggers (SRTTs) for a particular SRT. For example, infant-mother dyads will develop idiosyncratic stylised carry signals, e.g. a shoulder touch, that trigger responses, carrying behaviour, that would normally be triggered by overt carry request behaviour, climbing onto mother’s back. (Halina, Rossano, Tomasello; 2013).

It is tempting for us — advanced representation users that we are — to view SRTs (e.g. the presence of fire) as the referents of SRs, and the SRs themselves as internally distributed representations of those trigger vehicles. But this would be to make the same sort of mistake I described in Section III. SRTs are not themselves conceptual generalisations, of some object or situation type such as *fire*. If they were, this account would have begged the initial question of how we become information-users in the first place. Rather, they are situations in the world that effect some organismal response. At this level, there just is the response, no additional *understanding* or recognition of the relation between the trigger and the response. If we catalogued an organism’s responses over time, we would likely find that

clusters of trigger features co-relate with clusters of response features. But these cluster co-relations would not be an indication that the organism has become sensitive to anything *beyond the triggers themselves*, that is, to some higher-level property. In terms of the rabbit example, it is because foxes are all *foxes* that when a rabbit develops an SR to  $fox_1$  and then to  $fox_2$  and then to  $fox_3$ , that this will translate into similar SR responses to future foxes as well. There is no need here to assume some sort of internal representation of foxiness to explain this regular response behaviour.

Thus far, we have not yet accounted for R-activity, the sort of intentional (in both senses of that word) cognitive capacity we want to explain. But we have at least begun the task of grounding it in something that can carry the burden of representation. Organisms *use* symbols by exploiting relationships between types of situations in order to elicit responses in other organisms. The more innate this symbol use behaviour is, the more the relationship exploited will be a natural one; the more learned the symbol use is, the more the relationship exploited will be conventional, as a result of locally-learned associations. Initially perhaps, if we are trying to tell the story of human language development, gesture and sound triggers are closely sensorily linked to SRTs, in a way similar to flagging. Onomatopoeic words seem to work like this. The more adaptable an organism is, however, the more it will be capable of developing SRs to SRTTs that are not feature-related to the SRTs for which they are stand-ins. A human child, for example, must learn, through a great deal of repetition, to respond to the utterance “kæt” as she would to the presence of an actual cat (though perhaps with a less intense response). Once a human, or other adaptable entity, has a suite of such abstract SRTTs, they can be combined to create new SRTTs and those in turn can be further combined. The resultant possibilities of combinations, as we know, is combinatorily explosive. Critically, word manipulation — talking, reading, writing — is itself an act in the world and, as such, will trigger SRs in turn. Finally, we have SRs to SRTTs, not in virtue of their connection to some SRT, but in virtue of themselves, as direct triggers. This is the kind of reflexive activity paradigmatic of R-level cognition.

What sort of behaviour might we find in an SR to an SRTT? The original significance response, of course, e.g. the child responds to the utterance “kæt” in the way that it would to the presence of an actual cat, by some neural activity and, at the organismal level, perhaps by becoming excited. But, in addition, there might be a response to the SRTT itself, that is, to the utterance: perhaps it has always been the child’s mother who has uttered “kæt” in the past and its SR to the presence of mother is positive, so the utterance itself then elicits a deep, positive, emotional modality. Some excellent work has been done to begin laying down the theoretical framework within which we can better investigate these sorts of complex, dynamic processes. Terrence Deacon (2011), for example, has developed some new concepts to help us theorise about how entirely new kinds of processes can emerge out of complex, cyclic, dynamics of this sort: “the

orthograde/contragrade distinction may offer a useful way to reframe the emergence problem. Indeed, ultimately we will discover that a fundamental reversal of orthograde processes is a defining attribute of an emergent transition.” (p. 226) And, from the other side of the fence, as it were, Christiansen and Chater (2008) investigate the development of language itself as it ‘evolves’ alongside the users who create it.

As the beginning of an externalist account of our representational capacities, this sketch might seem to raise more questions than it answers. Looming large for me, someone who has a particular interest in linguistics and the philosophy of language, is the uneasy awareness that in treating language solely as a ‘tool for representing,’ as this account seems to do, I have swept under the carpet the vastly complex ways in which language is in fact used, something Wittgenstein (1997) so evocatively and eloquently demonstrated in his *Philosophical Investigations*. And the big questions of consciousness remain: is there something more to understanding the connection between an SRTT and SRT besides knowing how to exploit it? How do we explain the phenomenal aspects of thoughts and feelings? More generally, how does the language tool lead to self-consciousness? But here I have been concerned only to give a plausible externalist *grounding* of our representational capacities, not a full account of them. We must begin somewhere.

To sum up: R-activity is the cognitive capacity that cognitive science has been most actively investigating for the past 70 years. No surprise then that the computational metaphor has been such a motivating force in the field, anti-representational challenges notwithstanding. But those challenges are ignored at our peril, since no account is forthcoming unless they are taken seriously. The externalist approach I have described here does that by grounding the representational part of R-activity in organism-level engagement with its environment, not in its neural activity. Thus, Dreyfus is correct to insist that, “*being-in-the-world* is more basic than *thinking* and solving problems; that it is not representational at all” (Dreyfus, 1146), but wrong to ignore the important consequences our representation *tools* have on the ways in which we can be-in-our- world: the pen really is mightier than the sword.

## References

- Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: MIT Press.
- Christiansen, M. & Chater, N. (2008). Language as shaped by the brain." *Behav. and Brain Sciences*, 31.5, 489-509.
- Clark, A. (2015). Predicting Peace ... In T. Metzinger & J. M. Windt (Eds). *Open MIND*:7(R). Frankfurt: MIND.
- Clark, A. (2006a). Language, embodiment, and the cognitive niche. *Trends in Cog Sci*, 10 (8):370-374.
- Clark, A. (2006b). Material symbols. *Phil Psych*, Vol. 19, No. 3: 291–307.
- Clark, A. (2002). Skills, spills and the nature of mindful action. *Phenom and the Cognitive Sciences* 1: 385–387.
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge: MIT Press.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis* 58: 7-19.
- Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, 101: 401–431.
- Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Bio and Phil*, 22: 547–563.
- Deacon, T. *Incomplete Nature: How Mind Emerged from Matter*. New York: W.W. Norton & Company. 2011.
- Dreyfus, H. (2007). Why Heideggerian AI failed ... *AI*, 171: 1137–1160.
- Dretske, F. (1988). *Explaining Behavior*. Cambridge: MIT.
- Edelman, S. (2008). On the nature of minds .... *Journal of Exp & Theo Artificial Intelligence*, 20(3): 181–196.
- Freeman, W. (2000). *How brains make up their minds*. New York: Columbia University Press.
- Freeman, W., & Skarda, C. (1990). Representations: who needs them? In (Eds) J. McGaugh et al, *Brain Organization and Memory*....Oxford: OUP, pp. 375–380.
- Grice, H.P. (1989). *Studies in the Way of Words*. HUP.
- Halina, M., Rossano, F., Tomasello, M. (2013). The ontogenetic ritualization of bonobo gestures. *Anim Cogn*, 16:653–666.
- Keijzer, F. (1998). Doing without representations which specify what to do. *Phil Psych*, II(3), 269-302.
- Meneary, R. (2007). *Cog. Integration*. Palgrave Macmillan.
- Noë, A. (2010). Vision without representation. In (Eds) N. Gangopadhyay et al, *Perception, action, and consciousness*:.... New York: Oxford University Press.
- O'Regan, K. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav and Brain Sciences*, 24: 939–1031.
- Ramsey, W. (2003). Are receptors representations? *J of Exp. & Theo Artificial Intelligence*, 15:2, 125-141.
- Rowlands, M. 2006. *Body Language*. Cambridge, MA: MIT
- Salay, N. (2016). Representation: Problems and Solutions. In S. Ohlsson & R. Catrambone (Eds.), *Proc 37th Conf of COGS*. Austin, TX: Cognitive Science Society.
- Salay, N. (2014). Dress Rehearsals, Previews, and Encores: A New Account of Mental Rep. ... *Theoria*, 80:1, 84-97.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, New Series*, 185 (4157): 1124-1131.
- Van Gelder, T. (1995). What might cognition be, if not computation? *JPhil*, 92(7): 345-381.
- Wheeler, M., (2007). *Reconstructing the cognitive world: the next step*. Cambridge, MA: A Bradford Book.
- Wheeler, M., (2004). Is language the ultimate artefact? *Language Sciences*, 26, 693–715
- Wittgenstein, L., & Anscombe, G. E. M. (1997). *Philosophical investigations*. Oxford, UK: Blackwell.