The Journal of Mind and Behavior
Spring 2019, Volume 40, Number 2
Pages 121–146
ISSN 0271–0137

Learning How to Represent: An Associationist Account

Nancy A. Salay

Queen's University

The paper develops a positive account of the representational capacity of cognitive systems: simple, associationist learning mechanisms and an architecture that supports bootstrapping are sufficient conditions for becoming a representation user. In terms of the debates within the philosophy of mind, this paper offers a plausibility account of representation externalism, an alternative to the internalist, reductive models of intentionality that still play a leading role in the field. Although the central theme here is representation, methodologically this view complements embodied, enactivist approaches to studying cognition.

Keywords: representation, learning, enactivism

The computational/representational metaphor for mind is still the central guiding idea in cognitive science despite many insightful and well-founded rejections of it (Chemero, 2009; Clark and Toribio, 1994; Dreyfus, 2007; Freeman and Skarda, 1990; Keijzer, 1998; VanGelder, 1995). 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 seems to be no plausible solution: How does this representational 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 representational/computational at all; and, on the other hand, those unwilling to let those intuitions go — and this is still the majority — have tried to bridge the explanatory gap by pushing the metaphor into service at the level of implementation, the brain. But if this move is mistaken, and I will argue here that it must be, then we are building a house of cards: we need a broader theoretical framework that integrates reductive internalist analyses with

---

I would like to thank Raymond Russ for his cheerful editorial assistance with this paper. Correspondence concerning this article should be addressed to Nancy A. Salay, Ph.D., Department of Philosophy, Queen's University, Kingston, ON, K7L 3N6, Canada. Email: salay@queensu.ca

embodied, enactive insights. How do we *become* the representation-using beings we are? This is the foundational question on which we need to build.

More recently, a third way has begun to emerge, from extended mind theory. On this view, our cognitive capacities are partly constituted by the external scaffolding, "widewear" (Clark, 1997), that supports them. We develop algorithms that extend our computational capacities, long division for example, and we offload memory tasks onto our environment: a string around the finger reminds one to pick up the dry-cleaning; a to-do list keeps track of tasks; and Alzheimer's patients achieve independence with the aid of sticky notes. Language itself, Clark and others (Cimatti and Vallortigara, 2015; Deacon, 1998; Donald, 1991; Logan, 2007; Wheeler, 2004) have argued, is a cognitive niche construction, a tool that "transforms … problem spaces in ways that (when successful) aid thinking and reasoning" (Clark, 2006a, p. 370). On this view, the logic-like structure of language is a cognitive resource that complements the pattern-completing capacities of our brains, allowing us to formulate inferences, make plans, think about tasks in a linear fashion, and so on, all cognitive activities that a human without language could not accomplish.

For reasons that will become clear, hybrid approaches which are grounded in internal representations, even minimally (c.f. Clark, 2006a, 2006b; Rowlands, 2006; Wheeler, 2005), suffer the same problems that afflict the mainstream view: only a wholly externalist account will do. We owe our capacity for representation to our skillful use of language, itself a representation tool. This hypothesis offers up many intriguing research questions,[1] but before we can tackle these, we need to establish its possibility: How *could* we become representation users without ourselves being representors?

The Problem

Setting the Terrain

While the cognitive/non-cognitive divide might be fuzzy, there is much so-called cognitive activity, perhaps a majority of it, that does not necessitate representational description. The constraint "when we are at our cognitive best" was meant to carve off that small subset of activity that does make representational appeal. For example, a full account of the cognitive mechanisms that make it possible for me to imaginatively plan my trip for next week 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

[1] And there are many: What is the relationship between bare perception, conscious perception, and the language skill? Does the experience of self-consciousness emerge out of broad social linguistic practises? If yes, how? If words and concepts are external to us, what constitutes our understanding of a sentence or thought?

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 remain 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 have specific uses, I will stipulate a term here, "R-activity," in order to conveniently and unambiguously refer to just that subset of cognitive activity that necessitates representational descriptions in the sense described above.[2] To repeat, I am not here *making the case* that the representational/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 have just noted, 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 and Kahneman, 1974). Much of our waking life is not self-conscious, 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 computational-representational framework: we have feelings, moods, emotions, interoceptions, all of which seem best explained in terms of complex systems concepts.

Accounting for R-activity is an especially tricky business, as C.S. Peirce pointed out repeatedly in the course of his analyses, because representation is a triadic relation: for something to be a representation, it must stand in one sort of relation to the thing it represents and in another sort of relation to the "interpretant," roughly, the translation of the sign into the object. If I utter some words — "The oil is smoking!" — my audience reacts to the situation the words represent — the fact that the oil is overheating — rather than to the utterance itself. They understand that the words are functioning here as representations, stand-ins for some situation in the world. My cat, on the other hand, in not understanding this feature of the utterance, reacts to the utterance itself: her ears prick up, her head swivels around, and her body prepares itself for flight. Explaining R-activity, then, is giving a naturalistic account of this three-way relation, of explaining how cognitive agents develop the capacity to use some physical processes, for example utterances, not directly, but as stand-ins for other physical

---

[2] Note that the extensions of R-activity and what Clark has termed "representation-hungry" problems, (Clark and Toribio, 1994, p. 403) don't precisely overlap. Planning for the future is a paradigmatic case for Clark, but some cases of future planning (scrub jay food caching, for example) might be effected in non-symbolic ways. R-activity, on the other hand, involves explicit symbol use.

processes. There are various conceptions of *representation*, however, at play in the broader philosophical and cognitive science literature. As I shall argue, the brain activity that underwrites action and perception is not representational *in the same sense* as personal level R-activity is. We need to become clearer about these different senses of representation.

Fortunately, others have begun to do this work (Chemero, 2009; Egan, 2014; Hagueland, 2000; Markman and Dietrich, 2000; Ramsey, 2007). William Ramsey's analysis is particularly useful because his taxonomy of representation picks out the three[3] central divergent ways the concept is used in cognitive science: IO-representation; receptor representation; and, S-representation. (Ramsey, 2007)

An IO-representation system takes symbols — X's that are interpreted as Y's — as input and produces symbols as output. This sense rides on Peirce's "symbolic" notion of representation introduced above in which a representation is a triadic relation between an object, a sign of the object, and an interpretation/translation between the sign and the object. Things are never, in themselves, symbols, though we often speak in this way when we are being sloppy. We might say, "this token represents Fred" (in the context of a game), but the token plays a representational role only when the requisite relata are in place, namely, when there is some interpretational process under which we react to it as we would to Fred, as when, during the course of a game we might say, "Fred, you are going to overtake Sally next turn!" referring to the plastic playing piece on the board. Taken outside of its respective symbol framework, the token has no representational force at all. The interpretant part of the relation then, as we shall be understanding it henceforth, is itself an action, the taking of the one for the other. Thus, an IO-representation system just is a system capable of R-activity and giving an account of how this is possible without appeal to some lower level symbol use amounts to giving a naturalistic account of this capacity.

A receptor representation system is one in which some internal structure *X* functions to indicate distal condition *Y* for the system whole or for some subsystem of the whole. This sense invokes Peirce's "indexical" notion of representation — *X* represents *Y* by way of being a reliable indicator of *Y*. Examples here include natural signs such as smoke (indicating fire), fogged windows (indicating high humidity), and of course neural activity (indicating the distal cause of the activity). Dretske's (1988) thermostat example was offered up as an information-theoretic system, representational in this sense. On that analysis, the bi-metallic strip inside a thermostat *represents/indicates* the ambient temperature of the surrounding air because it is by virtue of the strip's physical properties, which nomically covary with the rise or fall in temperature, that it has its functional role, namely, to turn the furnace on when the temperature drops below a certain level and off when the

---

[3] I am excluding here the fourth kind of representation Ramsey identifies, tacit-representation, since it overlaps in various ways with receptor and *S*-representation.

temperature rises above it. In other words, it is in virtue of its information carrying properties that it has its functional role. In the next section I explain why treating internal states as representational in this sense is misguided.

Lastly, an *S*-representation system is one in which some internal structure *X* functions as a model of some distal condition *Y* for the system. This sense maps to Peirce's "iconic" representation and to John Haugeland's (1991, p. 62) more general observation that a structure functions as a representation when it "stands-in for" environmental features that are not reliably present — *X* represents *Y* in virtue of being isomorphic to *Y*. Examples here include maps, photographs, and mathematical functions. Note that Dretske's thermostat can be viewed as an *S*-representational system[4] (importantly different from its received information-theoretic analysis) since thermostats function to regulate temperature by way of internal states that stand-in for temperature thresholds.

The reductive internalist hope of the mainstream approach is that thin versions of representational concepts — receptor and *S*-representation respectively — can be systematically fleshed out to become the thick IO-representation concept requisite for R-activity. I turn now to a discussion of how this thin–thick transition is supposed to work and why it won't.

Why Reductive Internalism Fails

Neural activity is taken to indicate (receptor represent) some feature of the environment for an agent only if (1) the neural activity nomically co-varies with the presence/absence of the feature in question; and, (2) this co-variance relation explains the functional role of those neurons within the organism. Ramsey argues that this analysis fails to establish the representational role of neural activity because it fails to establish its informational role. Indication, Ramsey reminds us, is an entailment relation: in order for *X* to indicate *Y*, "we need the involvement of an inference-maker" (2007, p. 201), something capable of responding to what *X* indicates — the information that it carries — rather than to *X* directly. Without such an inferring agent, information does not play a role in the account and, consequently, there is no justification for viewing *X* as functioning to represent within the system. An example will help us see why such an inferring agent is not forthcoming.

The activity of a set of neurons in the vomeronasal organ (VNO) of many mammals nomically covaries with the presence of certain pheromones in the environment. The presence of these pheromones, in turn, nomically covaries with the presence of potential mates. Being sensitive to the presence of these pheromones, then, has survival value for the species because the presence of those pheromones is also nomically co-related with the presence of potential mating partners.

[4] As I shall argue, this is the correct sense in which it is a representational system at all.

In other words, it is because of the information relation between the presence of pheromones and the presence of potential mates that the VNO sensitivity developed at all. Thus, it is the co-variance relation between VNO neural activity and the presence or absence of potential mates that explains the functional role of those neurons for the organism. Ergo, VNO neural activity is representational; it functions to represent potential mates to the organism. 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 play an integral role in the explanation. If the account can be given without appeal to this information at all, information is not playing a theoretically relevant role in the mechanism (Ramsey, 2007). The question then, with respect to such information-theoretic accounts of neural activity, is this: Is information playing this kind of integral role in the brain?

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 pm each day, the exploiter must be capable of responding to this higher order relation. Individual neurons don't have this second-order capacity: they are embedded within a cause–effect web in which sub-micro particles are the primary players. Indeed, at the level of neural activity, an organism-level entity such as *potential mate* isn't even a some*thing*, being at far too high a level of organisation for it to show up at all. 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 are capable of inferring these information relations (notwithstanding the previous point), we would have posited something capable of using the information embedded in neural activity, but only at a greater cost: infinite homuncular regress. In order to explain the intentional capacity of those fore-neurons, how they are able to make a connection between one thing, a pattern of neural activity, and something else, 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 the 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 is being debated 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: as Egan argues, it is "an intentional gloss . . . that shows, in a perspicuous way, how the computational/mathematical theory manages to explain the intentionally-described explanandum with which we began and which it is the job of the theory to explain" (2014, p. 128). The impulse to take "the gloss" literally begins with the misguided idea that organisms use their own inner states at all; they don't, they are constituted by them. Indeed, at no point is an organism ever aware of the relationship between its own neural responses and the entailments they carry, in the way that we have just become aware of it in telling the evolutionary story above. Craver and Bechtel (2007) develop this point in their excellent paper in which they argue that the "mystery" of top–down causation results from a confusion of constituency relations with causal ones. Interlevel relations — those that are between entity wholes and their mechanistic components — cannot be causal because the players in a causal relation must, at a minimum, be spatiotemporally distinct from one another:

> Given the compositional relations between mechanisms and their components, the space–time path of the mechanism includes the space–time path of its components. They coexist with one another, and so there is no possibility of their coming to spatiotemporally intersect with 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. (Craver and Bechtel, 2007, p. 552)

Likewise, the increased motion of the molecules "inside" a bowl of soup does not cause the soup to heat up — the motion of molecules constitutes its heat; the cells "in" an animal do not cause it to exist — they constitute its existence. Describing these constitutive relations in a causal way is to create unwarranted and ultimately misleading dualisms.

Returning to the example, that there is a co-relation between the presence of certain pheromones and the presence of potential mates is plausibly a factor that played a role in the evolution of VNO sensitivity to those pheromones in the first place: organisms that had such sensitivities ended up mating successfully more often than organisms that did not and so those sensitivities got passed on. There is no need to advert to any information-use in telling this evolutionary account. Being sensitive to the presence of certain pheromones itself bestows a selectional advantage; neither the organism nor the organism's brain nor mother nature herself need *know* that the presence of those pheromones carries information about the presence of potential mates for the advantage to work. By supposing that neurons function to provide information to organism wholes, such accounts conflate evolutionary analyses of the development of traits and sensitivities, where it is appropriate to cite salient co-relations such as the one between pheromones and potential mates, with descriptions of straightforward causal activity, neural sensitivity to pheromones.

The skin on my body is sensitive to different forms of light: when strong ultraviolet 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 see them 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 information-theoretic accounts of some cell activity and mechanistic accounts of others. But teleological explanation is as unwarranted here as it is in scientific explanations more generally.

Our neural activity is, of course, a part of what makes it possible for us to engage in representational activity. But, for all of these reasons, it cannot bear the full explanatory burden of it. As Freeman and Skarda (1990) argue, only when we finally let go altogether of the idea that neural activity is representational will we begin to make headway in developing a theory of the brain:

> These considerations give an answer to our question about representations. Who needs them? Functionalist philosophers, computer scientists, and cognitive psychologists need them, often desperately, but physiologists do not, and those who wish to find and use biological brain algorithms should also avoid them. They are unnecessary for describing and understanding brain dynamics. They mislead by contributing the illusion that they add anything significant to our understanding of the brain. They impede further advances toward our goal of understanding brain function, because they deflect us from the hard problems of determining what neurons do and seduce us into concentrating instead on the relatively easy problems of determining what our computers can or might do. (p. 380)

The way in which neurons function as representations, then, cannot be as receptor representations. *S*-representation, on the other hand, looks to be a more promising theoretical construct, from the perspective of naturalism, since mindless systems can become *S*-representors.

Recall that an *S*-representation is any structure isomorphic to some distal condition that functions within a system to model that condition. *S*-representation is thus a dyadic relation — *X* represents *Y* just in case *X* is a model of *Y* and, importantly, no interpretation of *X* as *Y* is needed for *X* to fulfill this modelling role. A train track, for example, successfully guides a train from point A to point B precisely because it is a physical model of the route from point A to point B; if the track were not such a model, it couldn't serve as a guide. In this sense, the track *S*-represents the route. By virtue of syntactic features alone, then, *S*-representations are capable of playing semantic functional roles. This is the context in which Haugeland's slogan "Take care of the syntax and the semantics will take care of itself" (1981, p. 23) is correct. Recent predictive processing models of cognition (Clark, 2016; Friston, 2008), which make heavy use of the idea that brains function by way of models and hypotheses, might fruitfully be thought of in *S*-representational terms. On such a view, neural activity does not play an information theoretic role — as I argued earlier, this is a dead-end idea as far as representation is concerned — rather, in mirroring salient features of an organism's world, both concrete and abstract, neural activity *S*-represents those features. Resonating in this isomorphic way makes it possible for organisms to hook onto their environments in life-optimising ways: "That distributed inner model is itself the result of self-organizing dynamics operating at multiple temporal scales, and it functions selectively to expose the agent to the patterns of stimulation that it predicts. The generative model thus functions — just as an enactivist might insist — to enable and maintain structural couplings that serve our needs and that keep us viable" (Clark, 2016, p. 293). This is the non-question-begging sense of representation we need to be using whenever we advert to the representational capacities of brains.

S-representation, however, is still a long way from the IO-representation we ultimately need to explain if we are to give a comprehensive account of cognition. Non-cognitive systems, as I've noted, can be *S*-representation systems; consequently, while being an *S*-representational system might plausibly be a necessary condition for being a cognitive system, it cannot serve as a sufficient condition. Unfortunately, the gulf between dyadic *S*-representation and triadic IO-representation is easily overlooked. Even Clark, who clearly understands that they are distinct notions,[5] treats *S*-representational brain-level predictions, models, and hypotheses as though they were merely on a continuum with the personal-level IO-representation use constitutive of R-activity: "… the cognitive role of personal

---

[5] As we can see here: "Talk of information … must ultimately be cashed simply in terms of the energies impinging upon the sensory receptors. This is essential if we are to avoid, yet again, the illicit importation of an observer's perspective into our account of how informed observers are naturally possible in the first place" (Clark 2016, p.15).

narratives: the stories we tell, to ourselves and to others, about the flow and meaning of our lives. Such narratives function as high-level elements in the models that structure our own self-predictions, and thus inform our own future actions and choices" (2016, p. 286). The central problem is to explain how R-activity is possible; a part of that story might involve an account of the role that internal *S*-representations play in achieving this, but it will not be a scaling up account since a dyadic relation cannot scale up to a triadic one.

If brains alone cannot be the ground of R-activity, then we need to cast our theoretical net more broadly, as Clark (2006a, 2006b, 2015) and others (Donald, 1991; Jeffares, 2010; Logan, 2007; Menary, 2007; Rowlands, 2006; Sterelny, 2010; Sutton et al., 2010; Wheeler, 2004, 2005) have been advocating for some time now. As I've already noted, however, many (Clark, 2002, 2006a, 2006b; Rowlands, 2006; Wheeler, 2004, 2005) who recognise the inherently representational nature of what I am calling R-activity, but who are also compelled by the insights of embodied challenges to mainstream cognitivism, have been developing hybrid internalist representational accounts. While I am sympathetic with this motivation, I am not optimistic that the approach, to replace internal objective representations with internal action-oriented (or otherwise qualified) representations, will do for the reasons rehearsed here. I now show how an organism with a simple, associationist learning mechanism and an architecture that supports bootstrapping could develop a symbol tool.

Bootstrapping an Externalist Account

Significance Triads

Following in the tradition of stimulus–response theories of learning, I will take a stimulus to be something in an organism's environment that provokes an organism-level response, e.g., secretion of saliva, increase in heart rate, locomotion, perception. A stimulus is unconditioned (US) if it provokes an instinctive response, in other words, an unconditioned response (UR). A stimulus is conditioned (CS) if it provokes a response, a conditioned response (CR), only after a period of associative learning. Note that I will not follow the cognitivist trend in learning theory of treating neural responses to external stimuli as themselves stimuli. Treating neural activity as stimuli is to make the same interlevel/intralevel mistake discussed in the previous section. In what follows, then, it should be understood that stimuli and responses are all organism-level occurrences.

To use the now classic example of classical conditioning, Pavlov (1927) showed that an unconditioned stimulus–response relationship in dogs, between meat (US) and salivation (UR), could be exploited to establish a new stimulus–response pairing, between a previously neutral stimulus, the buzzing of an electric bell (CS), and the UR. After repeated experience of the constant conjunction of US/CS, the UR is elicited in situations in which the CS alone is present. Thus conditioned dogs thereafter respond to the buzzer as though to the presence of meat, by salivating.

In the laboratory, arbitrary associations of this sort serve as a useful control for innate responses. But in the wild, associations are seldom arbitrary in this way. Often it is the case that one bundle of features — bundled because they belong to some natural kind and so always appear together — is co-related with another bundle of features, perhaps because there is a causal relationship between the two. To use a well-worn example, fire (US — aversive for many animals) is almost always accompanied by smoke. As a result of experiencing the constant conjunction of the two, an organism capable of associative learning will, over time, develop a conditioned response to smoke. Such an organism, after learning, will respond to smoke as though to fire, even in the absence of fire.

This relation between US, CS, and UR/CR is what I shall be calling a "significance triad." So far, there is no representation here at all, neither internal nor external; however, there is significance, the seeds out of which representation will grow. When the organism responds to a CS as though to a US, the CS functions as a sign of the US for the organism, in the following sense: it is because fire poses a threat to an organism that it develops a response to it; but, it is because smoke (typically) accompanies fire that an organism develops a response to it. In other words, if fire weren't response-provoking, smoke wouldn't be either. Smoke, as Grice (1989) famously observed, is a natural *sign* of fire, but we can now see that this is a rather loose way of putting the matter: more perspicuously we should say that smoke is a natural sign of fire f*or organisms capable of learning the association between smoke and fire.*

But if a triadic relation has been established, a significatory one at that, haven't we got a proto-representational relation as well? Not quite. In some cases, once learning has occurred, the triad effectively collapses into a dyad: CR to CS. These are the cases in which, at the point of response, the relation between the original US and CS is no longer accessible. Classically conditioned ants, for example, are likely in this category: when faced with a CS, although there is an historical link to the US, this link is not accessible to the ant. Other animals — humans, great apes, and likely many others — on the other hand, have the capacity to access these past connections, to differing degrees. More on this below.

Thus, what makes the original triad a significance triad is the significatory connection between US and CS, but what precludes the triad from being representational is its instability: in those organisms that cannot maintain a connection to their own learning histories, the triad collapses into a dyad. Only when a triadic significance relation is stable, will it also be a relation of representation. The gulf between learning to respond to something as though to something else and doing so while also maintaining access to prior learning experience is wide indeed. Next, I discuss two of the conditions required to traverse this breach.

Two Dimensions of Learning

Thus far, I have been speaking of the learning of significance triads in a rather one-dimensional way, as though, once learned, it manifests in more or less the same way in all organisms. There are many important distinctions to be made with respect to learning, but here I want to focus on just two general dimensions relevant to this account, what I will be calling "flexibility"[6] and "fluidity."

Flexibility is the degree to which a system can learn new behaviour: the more and varied the possible new behaviour a system is capable of learning, the greater its flexibility; the fewer and less varied the new behaviour it is capable of learning, the less its flexibility. Examples of new behaviour include the following: a rat learning to press a lever; a human child learning to speak; a young chimpanzee learning to use a stick to "fish" for termites; a bird learning a novel call sound; an otter learning to use a stone as a hammer; a dolphin learning to spin; an artificial neural network (ANN) learning to respond to an input class with a new output. Flexibility is thus an attribute of a learning system. Completely inflexible systems are those that cannot learn new behaviour at all: a rock, a tornado, and an ANN that has achieved equilibrium are all examples of inflexible systems.

Some factors influence flexibility directly — brain size, neural plasticity, manipulable body parts, for example — while others do so indirectly, by influencing the motivation to learn new behaviour. Much could be said here on both scores, but I will leave these details to future fleshing out. The important point here is that an organism might be more or less capable of learning a new behaviour. Since learning to use representations constitutes a new behaviour on this account, being at the flexible end of this continuum will be a prerequisite.

Fluidity is the degree to which a system can access learned relations. It is an attribute of expectation: the more a system's expectations are multi-directional, the more fluid it is; the more a system's expectations are fixed in the direction in which they were learned, the less fluid it is. In predictive processing terms, we might think of fluidity as the degree to which a system combines multi-level, top–down predictions with bottom–up sensory processes: "Such systems exhibit powerful forms of learning and . . . are able flexibly to combine top–down and bottom–up flows of information within the multilayer cascade" (Clark, 2016, p. 26). Organisms implement fluidity in nervous systems. There are many factors that govern the proliferation/atrophy of synaptic pathways, but once connections are established, the activity of neurons at one end of a pathway will induce other neurons along the pathway to move to a state of high energy and thereby become primed to release their own action potentials whenever conditions exceed a certain threshhold or level of "expectation." Actual nervous systems are so vast

[6] Broadly, this is phenotypic plasticity, but I introduce a new term because I want to restrict the discussion to a specific sort of learning, namely the learning of new behaviour.

and complex that, although we have increasingly sophisticated tools for making detailed neural observations, interpreting the data is an ongoing challenge. Artificial neural networks are useful here because we can use them to model such synaptic "expectations." Learned response relations are encoded in networks as patterns of activations: when one node in such a network becomes active, the others to which it is connected become active to a degree dictated by the functions imposed by the model. In a recurrent network, stable patterns of behaviour can be learned and recalled from a small subset of the nodes engaged in its activation. Such systems are fluid on this view because activation dependencies are multi-directional — expectation of an output is heightened via activation of an input, but expectation of an input is heightened by activation of downstream, output nodes, as well. In feed-forward networks, on the other hand, though they are also capable of learning new behaviour, activation dependencies move forward only: expectation of an input would not be heightened by the activation of an output node in such a network because input activation is not affected by the activation of nodes "downstream." This sort of system has low fluidity. An example will make this attribute clearer.

Suppose that both meat powder and sucralose elicit a salivatory response from a given organism, O. In other words, meat powder and sucralose are unconditioned stimuli (US) for an unconditioned response (UR) in O. Call the two stimuli US_1 and US_2 respectively. If O is a fluid system, its exposure to US_1, and its consequent response of UR, will raise its expectation for US_2. We might say that, as a result of the occurrence of UR, O has become primed for US_2. We can test this by measuring O's response rates to US_2, both after UR and in its absence. If O responds to US_2 more quickly after UR than it does in its absence, we conclude that there has been a priming effect. A system that is not fluid, on the other hand, will show no such expectation effects. Because such low fluidity systems always move in a fixed way from stimulus to response, the UR should not prime O toward the US_2. No backward-looking, so to speak, expectations are built up in such systems.

At a phenomenological level of description, fluidity is manifested as the pull to do whatever can be done in a given situation, to use what is "ready-to-hand" (Heidegger, 1927/1962; Merleau–Ponty, 1945/1962). This heightened expectation is what "tunes" an organism to the possibilities in its environment, that makes it the case that, if a usable tool is present, that tool will be noticed. 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" (2007, p. 1158). The more expectations there are, the more there is in the environment that is thus ready-to-hand. The opposite is true as well: the fewer expectations there are, the fewer tools are available. This connection between expectation and what is ready-to-hand is nicely expressed by Maslow's

Hammer: "it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail" (Maslow, 1966, p.15). In other words, when one is expecting only one thing, the environment is replete with that one thing. An important consequence, then, of high fluidity is a widening of the scope for action.

Taken together, an organism that is both flexible and fluid will have a wide range of behaviour and will have a large array of possible actions from which to choose in any given situation. For such organisms, "decisions" become an important factor in action since there might be a relative stalemate in the internal/external pushes and pulls of a situation. Later we shall see how the language tool can serve as a vehicle for these decisions. Organisms that are at the low end of this attribute continuum, on the other hand, have a smaller range of action possibilities and, consequently, a relative lack of competition with which to contend: what is to be done next is, typically, clear. With these pieces in place, I turn at last to a discussion of the development of representation.

Proto-Representation

Thus far, I have not accounted for R-activity, the sort of intentional (in both senses of that word) cognitive capacity I am trying to explain. But I have begun the task of grounding it in something that can carry the burden of representation, namely, relations of significance, what I have been calling "significance triads." In learning a significance triad, an organism is capable of responding to a stimulus *as though to* another stimulus, e.g., Pavlov's dog responds to the buzzing bell *as though to* meat. There is a promising mirroring here of the representation relation, but, as we saw above, this is not yet representation. The response *as though to* could collapse into a simple dyadic relation if the original learning situation is no longer accessible to the organism: the dog responds (CR) to the buzzing bell (CS) but the *as though to* prong of the relation drops out of the picture. Thus, although there is an important learning history here, the response relation itself needs more in the way of occurrent scaffolding to serve as a representation relation. Flexibility and fluidity form the basis of this platform structure.

Flexibility is an obvious precondition. The more flexible a system is, the more relations it can learn. A system that has no flexibility has no capacity to learn any relations at all, and, consequently, cannot learn representation relations in particular. Without some degree of flexibility, then, an organism cannot learn triadic representation relations.

In highly fluid systems, expectation is raised in a multi-directional way throughout its web of learned relations. A fluid system that has learned to respond to smoke *as though to* fire will, on perceiving smoke, have a heightened expectation for all three elements of the significance triad, namely, (more) smoke (CS), fleeing (UR/CR), and the original stimulus itself, fire (US). Having the capacity to thus "access," as an expectation, the elements of these learned relations is a precondition to their being manipulable, a requirement if the elements are to be used in a significatory way. How might such significatory usage develop in a flexible and fluid organism?

Imagine an animal (*A*) standing on a ridge, looking down into a broad valley. *A* sees, in the distance, a large fire moving toward its family group, also in the valley. As the fire is downwind of them, the family in danger is completely unaware of the fire. *A* sees this and wants to warn them. (We do not need to assume mental activity in order to get this desire off the ground, as all social animals have an innate drive to warn their kin of impending trouble.) At the same time, *A* is also responding (UR) to the danger itself: adrenalin is pumping through *A*'s body, its pulse is rapid, its senses are highly tuned to its immediate surroundings. In other words, it is preparing for flight. As *A* is a flexible, fluid organism, let us suppose, it is also in a broad state of heightened expectation for a wide range of behaviour, in addition to its instinctive flight response. Experientially this is manifested as an attentional pulling toward various aspects of its environment, of certain features of *A*'s surrounding coming into sharp relief against others that fade into the background. As *A* turns and moves, the space of possibilities and, consequently, expectations/motivations is continually shifting: if a tool is approached or handled, the significance relations associated with it become heightened and other relations, that had been in expectation, diminish. If *A* then drops the tool or is distracted by a sharp sound, yet another set of relations becomes heightened. And so on in a fluid, dynamic, unfolding of the possibilities of the current situation.

Suppose that the remains of a fire smoulder nearby. *A* is drawn to it. Why? *A* has learned to respond to smoke *as though to* fire. Being in a state of expectation for all elements of this significance triad, smoke is a particular draw in the current emergency situation — it is ready-to-hand. *A* is thus drawn to pick up one of the smouldering pieces of wood and wave it overhead. Once *A* is moving and waving the burning log, the focus once again shifts and a new set of expectations is heightened. Thus is *A* drawn further along through the situation.

Let us stipulate here that *A* has no attendant thoughts — it is so "caught up" in the situation, so present, that conscious thinking, even were it possible, is not present. But even though, as I am supposing here, there is no attendant inner understanding, no conscious "I am using this to do this," *A*'s smoke-waving activity is nevertheless an example of representation use, albeit a rudimentary one: *A* is *using* smoke to evoke the response that fire generally triggers, and in this use, the smoke is *standing in for* fire itself. Here, at last, is the germ of the relation we need, one that doesn't just look like the representation relation if we squint, but one that actually instantiates it.

An obvious objection here is that something much more complicated must be going on, under the covers as it were, for *A* to act in the way it does. The fact that *A* responds to smoke *as though to* fire doesn't explain why *A* is drawn to wave a burning stick about in order to make others respond in a similar way. For that course of action to make sense, surely *A* needs to recognise (think/assume) that others have responses similar to its own; it must have theory of mind (TOM). Without this assumption, the premise, "*I* respond to smoke by fleeing" and the conclusion, "Therefore, others will respond to smoke by fleeing," remain unconnected.

As a piece of reasoning, this is certainly lacking. The disconnect here, however, is not a deficiency in rationality; rather, it is a misguided description of *A*'s actions as an instance of reasoning at all. Reasons and reason*ings* are made in the context of full-fledged representations. As I am still in the process of establishing, what we might call, "proto-representational" relations, whatever relations we use to describe the actions of our proto-representational *A*, they won't be *inference* relations.

Indeed, there is evidence that human children act in analogously irrational ways. When a child's vision is obstructed by something, say a blanket draped over its head, it acts as though nothing can see *it*: because it can't see, it can't be seen. As a piece of reasoning, this is poor. In order for the premise "I cannot see anyone" to lead to the conclusion "No-one can see me," there must be an intermediate premise something like "When one cannot see, one is likewise invisible to others" for which, being patently false, the child cannot have evidence. More plausible is to suppose that neither the child nor *A* is engaging in any conscious reasoning at all. *A* is drawn to create smoke because *A* responds to smoke and thus smoke is ready-to-hand for *A*; the child acts as though it can't be seen because it cannot see. In both cases the actions are unmediated responses.

Let us step back and see where we are now. First, I discussed how the natural conjunction of features in the world could result in learned responses to otherwise neutral stimuli, such as smoke. I called these the learning of significance triads, learning to respond with CR to CS *as though to* US, even in the absence of US. The capacity to learn these significance triads is what I have been calling flexibility, which is just a capacity to learn associations at all. As I noted, however, while having the capacity to learn significance triads is a prerequisite for intentionality, it alone is not sufficient for intentionality since an organism that is merely responding to CS *as though to* US is not thereby using CS as a sign for US. Indeed, though the CS/US pairing is critical during the learning process, once the relation is learned, the triadic relation can collapse into a dyadic one: CR to CS. Something more is required to keep the relation triadic and, thereby, a possible grounding for an IO-representation relation. Fluidity is this attribute. Fluid organisms retain far-reaching connections to their own learning history via the mechanism of expectation. In particular, they retain an expectation connection between CS and US: whenever the CS is experienced, there is a heightened expectation for the US as well, and vice versa. In phenomenological terms, these heightened expectations form what is ready-to-hand in the environment. When such a system is drawn to *use* what is ready to hand, to use CS in order to elicit a response to US, it is doing more than merely responding to the learned relation (in the dyadic way); it is using CS as a sign of US to elicit UR/CR.

Here, at last, we have a proto-representational action. But we do not have full representation yet. There is still a rather wide gulf to cross, between the sort of inflexible, hard-wired, significatory activity of animal calls and the flexible, learned, systematic, representational activity of animal language. More on this below. But it is important to see that the representational grounding for the more robustly representational activity that language affords is here in these simple, significatory activities, which is what one would expect from an account that is consistent with evolutionary biology. There is a continuum of representational activity, with the more innate, inflexible, unsystematic, alarm call activity at one end and the learned, flexible, systematic, language activity on the other. In the middle we find learned calls that admit of limited flexibility and systematicity, such as the regional calls of some bird species and the gestural systems of the great apes. It is tempting to view the activities at the beginning of this continuum as not representational at all, as collapsing into dyadic relations, since it is unclear whether organisms retain access to learning over evolutionary timescales or whether there is just the learned response and no subsequent heightened expectation for the original US. This is an empirical question that will need to be investigated. Either way, fluidity is thus a measure of the degree to which an activity on this continuum is representational. I turn now to the development of the language tool itself.

The Language Tool

As in the smoke/fire example I've been using, a sign vehicle might be some phenomenon that naturally co-occurs with an unconditioned stimulus or it might be some phenomenon that naturally co-occurs with an unconditioned response. Some animal alarm calls seem to work like this. Flagging, when an animal lifts its tail and displays the white fur beneath, serves to trigger the flight response in nearby animals, presumably because the white fur is also exposed during the flight response itself, as the troupe of animals is fleeing. Of course, flagging and many other animal alarm calls are often innately driven significatory actions and, as such, lie at the inflexible end of the representation spectrum.

The more sensorily distant a sign vehicle is from a stimulus or a response, the more explicitly it must be learned. Smoke is a feature that naturally occurs when fires are present and the white fur underneath a raised tail is a natural feature of a fleeing situation, but the more flexible an organism is, the more arbitrary sign vehicles it can learn. Bonobos, for example, have a wide repertoire of gestures and vocalisations, some of which seem to be grounded in the sort of natural relation described above, such as alarm calling, and others that seem to be locally learned, ontogenetic ritualisations that serve as sign vehicles. Infant–mother dyads, for example, will develop idiosyncratic, stylised carry signals — shoulder touches — that trigger responses — carrying behaviour — that would normally be triggered by overt carry-request behaviour — climbing onto mother's back (Halina, Rossano, and Tomasello, 2013).

Initially perhaps, if we are developing an account of human language development, gestural and vocal sign vehicles are closely linked, sensorily, to their objects; they are naturally co-occurring features of stimulus or response situations. But, as in the case of bonobos, increasingly abstract short-hands for these natural sign vehicles develop. The use of onomatopoeic words and stylised gestures, for example, takes us one step further along the representation continuum; increasingly arbitrary sign vehicles follow. A human child must learn, through a great deal of (often explicit) teaching/learning experience, to respond to the utterance "'kæt" as she would to the presence of an actual cat. Mother, pointing to the cat says, "Cat. Look at the *Cat*. Do you see the *Cat*? What colour is the *Cat*?" and so on. We now have good evidence that the more of this teaching/learning experience a child has with words, the more sophisticated her language comprehension and use will be; and, the stronger many other cognitive skills will be as well (Haak, Downer, and Reeve, 2012; Sénéchal and LeFevre, 2002; Suskind, 2015; Swanson, Orosco, and Lussier, 2015; Yeong, Fletcher, and Bayliss, 2017).

As has been observed (Chomsky, 1968; Davidson, 1975; Hockett, 1977), language can be rigorously distinguished from animal signaling of the sort I've been describing thus far by, at least, two key features: compositionality and productivity. Animal communication systems seem to be made up of units that can neither decompose nor be combined, or, if they can, only to limited degrees and in fixed ways. Likewise they do not seem to be productive, with the repertoire of calls for a given animal being more or less fixed for its lifetime. These observations support the internalist hypothesis that some innate factor, beyond flexibility and fluidity, must account for language development.

Hauser, Chomsky, and Fitch (2002) have suggested that an innate recursive capacity is required to explain our productive use of language. If a capacity for recursion entails a capacity for decomposition, for seeing the parts that make up a whole, some studies (Van Leeuwen, Verstijnen, and Hekkeit, 1999) seem to indicate that humans are lacking in this quarter: while we are adept at mentally combining images — imagine a horse with a horn — we are poor at mentally decomposing them — rotate the bottom right quadrant of an image you are remembering. The finding is that decomposition of this sort requires continual, sensorimotor interaction with the object under analysis. There is some evidence, then, that a capacity for recursion might not be an innate biological feature of the modern human. Perhaps the simpler capacity for combining alone could account for the compositionality and productivity of the language tool we have developed. We see a strong inclination to exercise this capacity in modern human children who seem compelled to stack blocks, pile sticks, and group objects — as well as in other animals: if there are objects to be moved and manipulated, an orangutan will unfailingly bring them together; birds and chimps stack twigs; and, octopi combine objects to solve problems.

But whichever internal factor it is that plays this crucial role in the development of language, we run the risk of becoming side-lined by reductive internalism, again, if we focus solely on a capacity-based explanation. A more comprehensive account will look at the subtle and gradual co-development of sign vehicle use, on the one hand, and, of the sign vehicles themselves, on the other. Consider: the more specific a symbol is, that is, the more it is tied to a particular response, the more constraints there will be for usage and, consequently, the less use it will be in combination with other sign vehicles. I can use smoke, as a sign vehicle, in a number of ways, but those ways are quite constrained toward signaling emergencies. The ease with which a sign vehicle can be produced must also play a role here. Producing smoke is a rather laborious process. If one wanted to combine this sign with another, also difficult to produce, combining the two would augment those difficulties, making combining less practicable and therefore less likely to occur. The fewer usage constraints there are and the easier sign vehicles are to produce, on the other hand, the more a user will be drawn to combine them. Words, as sign vehicles, are ideal in this regard (for humans).

Some excellent work has been done to begin laying down the theoretical framework within which we can better investigate the complex, dynamic processes that govern language use and language development. Deacon (2011), for example, has introduced some new concepts to help us theorise about how entirely new kinds of processes emerge out of complex, cyclic, dynamics of this sort. From the other side of the fence, as it were, Christiansen and Chater (2008) suggest that language itself should be seen as an "organism" that evolves in response to selectional pressures. And, to name just one more from a growing number, Jeffares (2010) argues that our capacity to develop cognitive-enhancing tools rests on the scaffolding provided by earlier cycles of more primitive tool use. The central point here is that we should be looking both internally *and* externally in order to explain the features of language and language users.

It is instructive here to look at the way in which other animals use sign vehicles. Consider a typical interaction with Kanzi,[7] a 37-year-old bonobo who has learned to respond to hundreds of words and can use lexigrams[8] flexibly in order to communicate with humans. Kanzi, being a full grown male bonobo, often has energy to burn. After a morning of word play in the research centre, he is restless and twitchy. There are many games that he likes to play, but one of his favourites is *ball*. Imagine Kanzi as he paces about the room. There is a heightened expectation for

[7] I draw, here, on Savage–Rumbaugh's (1990) descriptions of Kanzi.

[8] Lexigrams are visual symbols for words that are used in language instruction with non-verbal animals such as great apes.

ball as a result of his restlessness — he is a fluid system — and so the lexigram for *ball* is a ready-to-hand tool for him in this moment. He picks it out from the sheets of lexigrams that lie about the room, points at it, and looks around expectantly. When someone responds, perhaps by looking in his direction, perhaps by saying, "What would you like to do Kanzi?" new possibilities open up. Kanzi points again to the lexigram and then to his interlocutor, indicating that they should go and play ball. As with *A*, there is no need to suppose that Kanzi's lexigram use is an outward sign of some inner representational life, that Kanzi already has "in mind" the desire to play and merely uses the lexigrams to communicate this desire. On the externalist view being developed here, there are no inner representations to be communicated at all. Kanzi is drawn, according to myriad attentional pushes and pulls that are continually shaping, dissolving, and reshaping, to whatever is ready-to hand in his environment, itself a product of Kanzi's past learning experiences and what is currently winning out in the competition for attention. Here, it is restlessness, excessive energy, that dominate, and so Kanzi is drawn to the lexigram for *ball*. There is no plan here, no thought, "Now I have to find someone to play ball *with*;" rather, because balls are always thrown *to* someone or caught *from* someone, when Kanzi sees a person, he sees a potential playmate.

Now consider that these sign vehicles are tools, cognitive-enhancing tools to be sure, but, more basically, they are tools for effecting change in one's environment: by yelling "fire" I can induce panic in a group of people; by making a request, I can acquire food. Seen in this light, they are new additions to what is ready-to-hand. As one gains facility with these tools, as they become increasingly ready-to-hand, the more they open up the space of possibilities. With them one can plan for the future — as Kanzi does when he sets up the conditions for ball playing — and refer back to the past — as Kanzi does when he points to a bite on his arm and signs "Matata bite."

But if we have no internal, innate, capacity for representation at all, we might wonder, how could this sign vehicle skill ever develop into *conscious* R-activity, into thoughts and conscious perceptions? Giving a detailed answer to this question takes us well beyond the scope of this paper, but in the interest of completeness, I will describe, in broad strokes, how I see this part of the account unfolding.

As soon as Kanzi "says" "you ball me," the utterance is now a thing in the world. It is some situation to which we can react. One could nod in assent, one could begin kicking the ball, or, one could describe the situation: "*You* say, 'you ball me.'" Describing things is, in itself, a thing to *do*. This meta-activity can continue: "*You* say, 'you say, "you ball me".'" And so on. Of course, except in certain contexts (philosophical ones, perhaps), there isn't much motivation for continuing in this way: the pushes and pulls of our past experiences as they unfold in the present environment will generally out-motivate such meta-meta-activity. On the other hand, our language-rich environment, overflowing with sign vehicles, presents an enormous pull to engage in this descriptive, meta-activity. Indeed, though they are not visible — sign vehicles are not something we *see* (unless they are words on a page, of course) — the environment of a literate, adult human contains words more than it contains anything else. This is quite remarkable, particularly so when we consider how unaware we are of this singular fact. If the phenomenological picture of ourselves as beings-in-the-world is correct, then, when we augment it with a world that is replete with sign vehicles, the traditional Cartesian mind–body dualism comes into soft relief: we are beings who, most fundamentally, act; however, as it happens, most of our tools for action are tools of reason, namely, words.

In bare perception, there is just action, a perceiving of the world. There is no self-awareness, no consciousness at all. As we become skilled language users and as our landscape becomes filled with sign vehicles, increasingly we use them to describe, to ourselves, the situations we are in. These self-descriptions, what we typically call "conscious thoughts," are not new things-in-the-world: they are simply descriptions, and, consequently, not items to be added to our ontology. But the mistake is easily made. Consider Fred, admiring a red peony in his garden. In bare perception, he simply perceives the flower. He is not aware of the "content" of this perception, of what the perception is about, as something separate from this flower in front of him. But now Sally sidles alongside and says, "I wonder what that red is like for you. For me it is so firey!" Fred frowns, perplexed. Sally, we might say, has just used some words to describe *her perception of the peony*. When we put the matter like this, it seems as though we now have something extra here, Sally's experience of her perception, what we sometimes call the qualia of her perception. But we don't have anything extra here at all. The thought that we do is a consequence of loose or sloppy describing. Sally has used some words to describe her description of the peony, not her perception of the peony. In bare perception, Sally is simply perceiving the flower as Fred is; she has no awareness of this at all. In conscious perception, however, Sally is using words to describe the situation — the peony perceiving — to herself. This is a description. We often call this description the perception itself, but on the externalist account being developed here, we distinguish between perceptions and descriptions of perceptions. When Sally uses words to report on this description to Fred, she is describing the *description*, not the perception.

Because descriptions lend themselves to reification (we can describe our descriptions), they lead to this sort of dualistic mistake, of supposing that there is some experience in addition to the bare perception. Descriptions mislead us in a second way as well. Since we must use the sign vehicles we have on hand, our descriptions will be shaped and constrained by the words we happen to have. The concepts we develop out of these descriptions, *self, personality, mind*, for example, are as much a function of the words we have as they are of the world itself.

On this view, then, it is the compounding effect of sign vehicle use and the proliferation of the vehicles themselves from which conscious thought — self-descriptions — emerges. This position is clearly sympathetic with those who see qualia talk as fundamentally misguided (Churchland, 1985; Dennett, 2001), but, unlike those, it offers an explanation for why thoughts, feelings, and emotions seem to have a separate existence. Thus, while this externalist view is firmly grounded in materialism, it also offers an explanation of the psychological draw toward dualism, something most reductive internalist views fail to address.

To summarise, then, when a flexible/fluid system develops a capacity for using sign vehicles, first in the very rudimentary way in which animals use distress calls to warn of danger, and then, in more sophisticated ways, using increasingly general and arbitrary sign vehicles to perform increasingly abstract actions, the beginning of R-activity emerges. The dynamic interchange between representational tools that are easily produced and combined with some internal capacity/inclination to combine them underwrites an explosive proliferation of vehicle and vehicle use. Over time, the world into which new humans are born becomes a language-rich one. A child's development into a literate adult is thus a symbiosis between its own flexibility and fluidity and the word-filled landscape in which it grows. How this skill with words eventually develops into thought and self-consciousness, what understanding amounts to, what distinguishes propositional knowledge from skills are questions still to be addressed, though I have given an idea of the direction in which the explanation will lead.

Conclusion

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 (1953/1967) 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 CS and US/UR 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? 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/representational 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" (2007, p. 1146) but wrong to ignore the important consequences our representational tools have on the ways in which we can be-in-our-world. Marshall McLuhan's (1964) insight, that the medium is the message, promises to take on even deeper meaning as we explore this new role for language.

References

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, Massachusetts: MIT Press.
Chomsky, N. (1968). *Language and mind*. New York: Harcourt, Brace and World.
Christiansen, M., and Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509.
Churchland, P. (1985). Reduction, qualia, and the direct introspection of brain states. *The Journal of Philosophy, 82*(1), 8–28.
Cimatti, F., and Vallortigara, G. (2015). So little brain, so much mind. Intelligence and behaviour in non human animals. *Reti, saperi, linguaggi, 2*(1), 5–22.
Clark, A. (1997). *Being there: Putting brain, body, and world together again.* Cambridge, Massachusetts: MIT Press.
Clark, A. (2002). Skills, spills and the nature of mindful action. *Phenomenology and the Cognitive Sciences 1*, 385–387.
Clark, A. (2006a). Language, embodiment, and the cognitive niche. *Trends in Cognitive Science*, *10*(8), 370–374.
Clark, A. (2006b). Material symbols. *Philosophical Psychology, 19*(3), 291–307.
Clark, A. (2015). Predicting peace: The end of the representation wars. In T. Metzinger and J. M.Windt (Eds.), *Open MIND*:*7*(R). Frankfurt: MIND.
Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.
Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis, 58*, 7–19.
Clark, A., and Toribio, J. (1994). Doing without representing? *Synthese, 101*, 401–431
Craver, C., and Bechtel, W. (2007). Top–down causation without top–down causes. *Biology and Philosophy*, *22*, 547–563.
Davidson, D. (1975). Thought and talk. In S. Guttenplan (Ed.), *Mind and language* (pp. 7–23). Oxford: Oxford University Press.
Deacon, T. (1998). *The symbolic species: The co-evolution of language and the brain.* New York: WW Norton.
Deacon, T. (2011). *Incomplete nature: How mind emerged from matter.* New York: W.W. Norton.
Dennett, D. (2001). Consciousness: How much is that in real money? In R. Gregory (Ed.), *Oxford companion to the mind*. Oxford: Oxford University Press.
Donald, M. (1991). *Origins of the modern mind*. Cambridge, Massachusetts: Harvard University Press.
Dretske, F. (1988). *Explaining behavior*. Cambridge, Massachusetts: MIT Press.
Dreyfus, H. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Artificial Intelligence, 171*, 1137–1160.
Edelman, S. (2008). On the nature of minds, or: Truth and consequences. *Journal of Experimental and Theoretical Artificial Intelligence*, *20*(3), 181–196.
Egan, F. (2014). How to think about mental content. *Philosophical Studies*, *170*, 115–135.
Freeman, W. (2000). *How brains make up their minds*. New York: Columbia University Press.
Freeman, W., and Skarda, C. (1990). Representations: Who needs them? In J. McGaugh et al (Eds.), *Brain organization and memory* (pp. 375–380). Oxford: Oxford University Press.
Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11).
Grice, H.P. (1989). *Studies in the way of words*. Cambridge, Massachusetts: Harvard University Press.
Haak, J., Downer, J., and Reeve, R. (2012). Home literacy exposure and early language and literacy skills in children who struggle with behavior and attention problems. *Early Education & Development, 23*(5), 728–747.

Halina, M.. Rossano. F., and Tomasello, M. (2013). The ontogenetic ritualization of bonobo gestures. *Animal Cognition*, *16*, 653–666.
Haugeland, J. (1981). *Mind design: Philosophy, psychology, artificial intelligence*. Mongtomery: Bradford Books.
Haugeland, J. (1991). Representational genera. In W. Ramsey, S. Stich, and D. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 61–89). Hillsdale: Lawrence Erlbaum Associates.
Haugeland, J. (2000). *Having thought*. Cambridge, Massachusetts: Harvard University Press.
Hauser, M., Chomsky, N., and Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569–1579.
Heidegger, M. (1962). *Being and time* [J. Macquarrie and E. Robinson, Trans.]. New York: Harper & Row. (Original work published 1927)
Hockett, C. (1977). *The view from language: Selected essays*. Athens: University of Georgia Press.
Jeffares, B. (2010). The co-evolution of tools and minds: Cognition and material culture in the hominin lineage. *Phenomenology and the Cognitive Sciences, 9*, 503–520.
Keijzer, F. (1998). Doing without representations which specify what to do. *Philosophical Psychology*, *II*(3), 269–302.
Logan, R. (2007). *The extended mind: The emergence of language, the human mind, and culture*. Toronto: University of Toronto Press.
McLuhan, M. (1964). The medium is the message. In *Understanding media: The extensions of man* (pp. 129–138). New York: Signet Press.
Markman, A. B., and Dietrich, E. (2000). In defense of representation. *Cognitive Psychology*, *40*, 138–171.
Maslow, A. (1966). *The psychology of science*: *A reconnaissance*. New York: Harper & Row.
Menary, R. (2007). *Cognitive integration*. Basingstoke: Palgrave Macmillan.
Merleau–Ponty, M. (1962). *Phenomenology of perception* [C. Smith, Trans.]. London: Routledge and Kegan Paul. (Original work published 1945)
Noë, A. (2010). Vision without representation. In N. Gangopadhyay, M. Madary, and F. Spicer (Eds.), *Perception, action, and consciousness: Sensorimotor dynamics and two visual systems* (pp. 245–256). New York: Oxford University Press.
O'Regan, K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24,* 939–1031.
Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* [G. V. Anrep, Trans.]. London: Oxford University Press.
Peirce, C. S. (1998). *The essential Peirce: Selected philosophical writings* (Vol. 2). Bloomington: Indiana University Press.
Ramsey, W. (2003). Are receptors representations? *Journal of Experimental and Theoretical Artificial Intelligence*, *15*(2), 125–141.
Ramsey, W. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.
Rowlands, M. (2006). *Body language*. Cambridge, Massachusetts: MIT Press.
Savage–Rumbaugh, S. (1990). Language acquisition in a nonhuman species: Implications for the innateness debate. *Developmental Psychobiology, 23*(7), 599–620.
Sénéchal, M., and LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2*),* 445–460.
Sterelny, K. (2010). Minds: Extended or scaffolded? *Phenomenology and the Cognitive Sciences*, *9*(4), 465–481.
Suskind, D. (2015). *Thirty million words: Building a child's brain*. New York City: Dutton Books.
Sutton, J., Harris, C. B., Keil, P. G., and Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences, 9*(4), 521–560.
Swanson, H., Orosco, M., and Lussier, C. (2015). Growth in literacy, cognition, and working memory in English language learners. *Journal of Experimental Child Psychology, 132*, 155–188.
Tversky, A., and 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? *The Journal of Philosophy, 92*(7), 345–381.
Van Leeuwen, C., Verstijnen, I., and Hekkeit, P. (1999). Common unconscious dynamics underlie uncommon conscious effects: A case study in the iterative nature of perception and creation. In J.S. Jordan (Ed.), *Modelling consciousness across the disciplines* (pp. 179–218). Lanhan, Maryland: University Press of America.

Wheeler, M., (2004). Is language the ultimate artefact? *Language Sciences*, *26,* 693–715.
Wheeler, M., (2005). *Reconstructing the cognitive world: The next step*. Cambridge: Bradford Press.
Wittgenstein, L. (1967). *Philosophical investigations* [G. E. Anscombe, Trans.]. Oxford: Blackwell. (Original work published 1953)
Yeong, S., Fletcher, J., and Bayliss, D. (2017). Impact of early home language exposure on phonological and orthographic skills and their contributions to English literacy abilities in English monolingual and Chinese–English bilingual adults. *Applied Psycholinguistics, 38*(1), 181–210.