Proceedings of the Annual Meeting of the Cognitive Science Society

Why Neurons Cannot be Detectors: Shifting Paradigms from Sherlock Holmes to Elvis Presley?

Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN

1069-7977

Salay, Nancy

Publication Date

2011

Peer reviewed

Why Neurons Cannot be Detectors: Shifting Paradigms from Sherlock Holmes to Elvis Presley?

Nancy A. Salay (salay@queensu.ca)
Department of Philosophy/School of Computing, Watson Hall 310
Kingston, ON, K7L 3N6 CANADA

Abstract

The practice of treating neurons as detectors is ubiquitous in the neuro-science community and in AI as well, in the context of neural networks. But there are a growing number of cognitive scientists who think that the representational paradigm is ill-suited to this level of explanation. In this paper, I rehearse William Ramsey's powerful critique of neural-detector attribution, focusing on his argument that Dretske-style information theoretic accounts of representation fail to justify the practice. I then take this conclusion a step further by arguing that not only does this particular justification fail, none at all are possible. The conclusion that we need to let go of the representational paradigm is not a negative one though, I shall claim, because it liberates us from the kind of misguided thinking that leads to theoretical dead-ends. Once we see this, we are free to investigate new, more fruitful, paradigms.

Keywords: representation; neurons; detectors; information theory; Dretske; interaction theory.

Introduction

The practice of treating neurons as detectors is ubiquitous in the neuro-science community and in AI as well, in the context of neural networks. But there are a growing number of cognitive scientists who think that the representational paradigm is ill-suited to this level of explanation. With this paper, I will add my voice to these ranks. In section one, I will rehearse William Ramsey's powerful critique of neural-detector attribution, focusing on his argument that Dretske-style information theoretic accounts of representation fail to justify the practice. In section two, I take this conclusion a step further by arguing that not only does this particular justification fail, none at all are possible. The conclusion that we need to let go of the representational paradigm is not a negative one though, I shall claim, because it liberates us from the kind of misguided thinking that leads to theoretical dead-ends. Once we see this, we are free to investigate new, more fruitful, paradigms. In the final section, I briefly discuss one of the more promising ones.

Section I: Ramsey's Critique

By explicitly addressing the question of what justifies neural-level detector attributions, Ramsey brings to the fore a discussion sorely lacking in the cognitive science community. I present here just two examples of this practice in order to focus this discussion[1]:

> 1. "… the key claim of localist coding schemes is that a given unit (neuron) codes for one familiar thing (and does not directly contribute to the representation of anything else), and that it is possible to interpret the output of a single unit in a neural network." (Bowers, 2009, p. 223)
> 2. "These investigators report the discovery of number-encoding neurons in the lateral prefrontal cortex of the macaque brain. ... this work opens up the exciting possibility of studying the cerebral bases of elementary arithmetic at the single-cell level." (Dehaene, 2002, p. 1652)

Unfortunately, because the practice is so ubiquitous and unchallenged in the field of neuroscience, there aren't many actual justifications to assess. As Ramsey points out,

> ... researchers often skip the question of whether neural receptors function as representations and instead ask about how the representational encoding is done. That is, researchers often begin with the assumption that neurons function as representations, and then explore, for example, whether the encoding is in single cell 'grandmother' representations or instead distributed across a population of neurons. (2003, p. 127)

I won't rehearse here Ramsey's speculations for why there has been so little written about this central topic, although his diagnosis is spot on; instead, I will skip to his decision to focus the critique on Fred Dretske's information theoretic account of representation, since, he concludes, this is the most robust, well-defended account that comes the closest to offering an explicit justification for the practice:

> ... his theory seems clearly motivated by examples of the very notion of representation we are trying to explicate, and many have appealed to Dretske as a way of defending receptor-style representations. What is more, because Dretske's account of content is so closely intertwined with an account of what it is for something to function as a representation, we see that he is, indeed, worried about providing a solution to what I have been calling the functional specification challenge. Thus, if anybody has given a carefully worked-out philosophical explication and defence of the receptor notion ... it is Dretske. (*ibid.*, p. 131)

---

[1] I do not intend to single these out as particularly egregious cases or so on in any way – there are literally hundreds of others I could have chosen, but a choice had to be made.

On Dretske's account, what makes some internal state X a primitive representation or detector of some class of things or actions Y is that it meets the following three conditions:

> 1. The presence/absence of X covaries with the presence/absence of members of Y;
>
> 2. The co-variance is under-written by a nomic causal relation, that is, the presence/absence of members of Y cause or are a necessary part of the cause of the presence/absence of X; and,
>
> 3. The functional role of X, within the system within which it arises, is to carry information about the presence/absence of members of Y. (Dretske, 1988)

Condition 3 ultimately does the work of justifying our treatment of X as a representation, since lots of states meet both conditions 1 and 2 alone, but do not function to carry information about and, consequently, represent anything. For example, the presence of large electrical fields is causally necessary for the presence of lightning, but lightning does not represent electrical fields. Now the trick, of course, to developing a fully naturalistic account of representation, is to explain how condition 3 can come about without appealing to the existence of some intentional system in which X functions to carry information. Dretske follows teleological-functionalists such as Millikan in arguing that such functional roles are established as a result of natural selection or, in some contexts, in the course of the development of learning mechanisms. Here is an excerpt from Dretske on how he sees such functional information-carrying roles being established:

> Suppose an animal – call it Buster – is so wired that it can see nearby *O*s.... Because *O*s are dangerous to animals like Buster, it quickly learns to avoid them. Learning to avoid *O*s is a process in which an internal sign of *O*, an internal signal carrying the information that an *O* is present, is made into a cause (a triggering cause) of whatever movements constitute avoidance. ... As a result of the learning of the sort just described, Buster's internal circuitry has been reconfigured so as to give an information-bearing element a control function." (1994, p. 69)

According to Dretske, it is in virtue of the information carried by the causal co-relation between the presence of *O*s and the internal *O* signal that the internal *O* signal gets its role. Here is another of Dretske's examples, one that Ramsey highlights, of the same sort of process, but one that develops as a result of evolutionary pressures:

> … the magnetesomes in anaerobic bacteria indicate the direction of magnetic North, which also happens to correlate with deeper, anaerobic water. Through a process of natural selection, these magnetesomes come to be wired to the bacteria's navigational system because of their nomic link to anaerobic water. They are thus given the functional role of indicating the direction of anaerobic water and, according to Dretske, thereby become anaerobic water representations. (Ramsey, 2003, p. 132)

Again, the idea is that it is in virtue of the information carried by the causal co-relation between magnetesomes and anaerobic water that, through natural selection, magnetesomes developed the functional role they did, within the context of anaerobic bacteria. This, according to Dretske, is what justifies our treatment of them as anaerobic water detectors.

But, Ramsey argues, this is much too quick. How do we know that it is in virtue of the *information* the causal relations carry that they were selected for? In order to get to that conclusion, Ramsey argues, we need a much more ontologically-loaded notion of information than is warranted, one in which information itself can play a causal role: "… many writers—including Dretske—appear to reify information with expressions like 'information flow' and 'information carrying' …."(2003, p. 135) But we have no independent justification for treating information in this way. From an ontologically Spartan vantage point, information is just what can be learned about the causal history of some object or system: "Talk about information carrying can be understood as simply a way of saying that nomic relations between states of affairs allows us to use these states of affairs to discover things." (Ramsey, 2003, p. 135) Indeed, as Ramsey points out, these states of affairs need not even be directly causally related to one another in order for there to exist an information relation between them. If A is larger than B, and B is larger than C, then A 'carries' information about C, since knowing something about A, say that it has length X, allows one to deduce something about C, say that it has length < X. In other words, "being an information carrier is nothing more than being a thing that stands in some sort of relation to something else, such that the former can be exploited to gain knowledge about the latter." (2003, p. 135) In this unreified sense, information abounds.

But although information abounds, it doesn't follow that all or indeed any of this information is in fact used. Two things might be causally related, let's say the presence of A causes some process to occur in B, but the fact that B's activity carries information about the presence of A may play no role at all in this causal transaction, not even in an account of the evolutionary history of the development of this causal relationship. For example, if I squirt a drop of water onto a small sample of salt, the salt will begin to dissolve. The salt's activity, the dissolving, is an indication that a liquid is present, that is, I could discover from its present state that a liquid is present and I could also discover, with the right equipment, exactly when in the history of this sample the liquid was introduced; but, of course, the dissolving will continue on whether or not I actually attempt to deduce this information.

Likewise, Ramsey points out, in the examples Dretske uses to support his case, in none of them is it clear that it is in virtue of the *information* that the underlying physical causal relations carry that it is selected for:

> For instance, the iron deposits that serve as magnetesomes in anaerobic bacteria are wired to the

> bacteria's propulsion devices because of the way they reliably respond to anaerobic conditions. We need some further reason, however, for thinking they are recruited into service because of the information that results from this relation. There is really no sense in which the bacteria's flagellum (their propellers) exploit the informational content carried by the magnetesomes: no sense in which they use the magnetosomes to discover something about anaerobic conditions. It is one thing to serve as a causal mediator between A (anaerobic conditions) and B (directional propulsion), it is an entirely different thing to serve as an informer about A for B. (2003, p. 137)

Ultimately, Dretske's account fails, then, because of the untenable, but critical for his view, distinction between the physical and informational features of causal relations; to get the teleological story off the ground, the informational relations need to play a causal role in the account. At the least, this assumption is as non-naturalistic as the very notion of intentionality it was invoked to demystify. At the worst, it is an ontological load too heavy to bear. Without it, however, the support for condition 3 is removed and, unless condition 3 is met, we aren't justified in treating causal relations as representation relations. We're back at square one.

Section II: No Justifications are Forthcoming.

In this section, I want to argue for the following stronger claim: not only do Dretske-style accounts fail to justify the practice of neural-detector attribution, but no such justification is in the cards at all.

Now I'm certainly not making a novel claim when I say that, conceptually-speaking, *representation* and *neuron* are concepts appropriate to different levels of explanation: within the cognitive science community, David Marr's tri-level hypothesis[2] has been widely accepted and used to justify division of labour[3]. In the context of this sort of level distinction, we could say that using the concept of a representation, which is a concept proper to either the computational or the algorithmic level of explanation, in order to pick out kinds at the implementation level – this is what we are doing after all when we treat neurons as detectors – is just to confuse levels of explanation. In order to perform this kind of reduction, we need a theory that allows us to bridge between the levels. This theory will explain *how* the more abstract, higher-level concept of representation is instantiated at the neural level.

Of course, finding strong co-relations between neural activity and states of affairs in the world in conjunction with a solid information theoretic account of representation is supposed to play exactly this bridging role. Indeed, it's because of a sensitivity to this abstractness of the concept of representation that researchers are typically careful to call neurons detectors, *primitive* representations, rather than full-blown ones. But, as we saw in the previous section, Ramsey's arguments undermine the justificatory support that information theoretic accounts give to treating neurons as detectors. Consequently, there is reason to be suspicious of the current scaffolding holding together the neural-representational hierarchy. My aim here is not to critique this Marr-inspired levels approach to cognitive inquiry, (although I do think it biases us towards a particular view of what could count as a cognitive process); rather, I want to accept this way of dividing the theoretical labour and argue that a further explanatory distinction we ought to make serves to limit the kinds of concepts we can use to theorise at the various levels. As a consequence, we will see that concepts such as *detector* can never be applied at the implementation level, no matter how much bridging we do.

To begin the deconstruction, we need to introduce a new kind of distinction, one that tracks the degree of context a given concept includes. Being a distinction of degree, we shouldn't expect too many instances at either end of the continuum; most concepts will fall somewhere along the middle, perhaps closer to one side or the other, of what I'll be calling the individual/collective continuum to indicate concepts that pick out kinds in virtue of their context-free features, on the one hand, and concepts that pick out kinds in virtue of their context-dependent features on the other. What it is to be an instance of a strongly individual concept will depend mostly upon the local, non-relational, properties its instances have. The concept *hydrogen*, for example, is highly individual, in this sense, because to be an instance of it is to meet a set of conditions that can be specified in a generally context-free way, e.g. being an atom with one proton in its nucleus. What it is to be an instance of a strongly collective concept, on the other hand, will depend mostly upon the system-level, relational, properties its instances have. For example, the concept *worker ant* lies closer to the collective side of the continuum since, while there are certainly some individual features that worker ants exhibit, e.g. being female, having a certain body size, and so on, it is not possible for an ant to be a worker ant unless there is an ant colony within which it can function in that way; a lone ant, outside of its colony context, is no longer a worker ant, since part of what it is to be a worker ant is to play a certain role within a larger system. Thus, certain concepts can be applied to individuals without appeal to the broader system within which those individuals are found,

---

[2] Marr (1982) describes a framework for the theoretical task of explaining visual processing, which we can extend to cognition in general, in which the following three levels of explanation are distinguished: at the highest level of abstraction, the computational level, we describe the general function of the system under investigation; at a middle level, the algorithmic level, we describe the processes or mechanisms that make this activity possible; and, at the lowest level, the level of implementation, we describe how the 'hardware' performs these actions.

[3] Griffiths et al., (2010) for example, are quite explicit that their theories apply to the function level of explanation only: " ... probabilistic models of cognition pursue a top-down or 'function-first' strategy, beginning with abstract principles that allow agents to solve problems posed by the world – the functions that minds perform – and then attempting to reduce these principles to psychological and neural processes." (Griffiths et. al, 2010, p. 357)

while others cannot be so applied — they necessarily involve some relational attributes.

Now, the concept of being a detector is clearly a collective concept. To see this, consider the following example. We might want to call a magnetised metal rod a metal *detector* in virtue of the causal relations that exist between it and instances of metal – metallic objects within a certain distance will, quite literally, be drawn towards the rod. But, as we saw in the previous section, such a rod is no more a metal *detector* than the magnetesomes in anaerobic bacteria are anaerobic water detectors. A magnetised metal rod can only have the functional role to detect metals within a context within which it is used in this capacity. This is because to be a detector is to play a particular role in a system, namely, to carry information about the presence/absence of members of a certain class. To notice this is just to acknowledge that there are certain features of the concept of being a detector that cannot be explicated by appeal to the purely individual features of an object acting in this capacity, since it's the playing of a certain role, and this is a relational attribute, that is essential to being an instance of the concept.

A final distinction will help tie this discussion back to levels of explanation. Andy Clark (1996)[4] convincingly argues that, in cognitive science, we ought to be distinguishing between three different classes of explanations, where each is differentiated according to how much context is included in it. For the sake of symmetry and because I don't want to get side-tracked here by controversies over emergence, I will ignore Clark's third category of emergent explanation and focus only on the first two: homuncular and interactive explanations.

We provide an homuncular explanation when we theorise about an individual by "adverting to the capacities and roles of its components, and the way they interrelate." (*ibid.* p. 5) For example, when we describe how a machine works by appealing to its sub-components, we are giving a homuncular explanation of it. We provide an interactive explanation when we include the role of the environment in our account of how some system functions in that environment. Clark cites Ballard's approach to understanding vision as an animate process as a good example of interactive explanation. In contrast to the traditional homuncular treatment of vision "as the task of building a detailed representation of a 3D world on the basis of what is essentially a body of 2D data," (*ibid.* p. 7) Ballard

> depicts the goal of vision as the production of successful actions within an environment context, keeping computational costs as low as possible. ... Thus, according to Ballard, the idea of a component which encodes a full-scale model of our surroundings is misguided. Animate vision, Ballard argues, neither needs nor can afford to create and sustain such a model. Instead, we constantly saccade around, picking up only such fragments of information as we need to support specific actions, and re-visiting the scene again and again rather than relying on some internally represented surrogate. (*ibid.*, p. 8-9)

Clark's context-based distinction between explanations complements the dichotomy between individual and collective concepts I have been developing: individual-level concepts are best explicated with homuncular explanations, while more collective concepts can only be fully characterised with interactive explanations, since only the latter will draw the relevant aspects of context into the description.

An example will help make clear how I see the homuncular/interactive and the individual/collective divides working together in explanations. Take the concept of an automobile. As with many concepts, there are both individual and collective aspects to it. From *Wikipedia*, for example, we get this definition:

> An automobile, motor car, or car is a wheeled motor vehicle used for transporting passengers, which also carries its own engine or motor. Most definitions of the term specify that automobiles are designed to run primarily on roads, to have seating for one to eight people, to typically have four wheels, and to be constructed principally for the transport of people rather than goods. (http://www.wikipedia.org/)

If we focus on defining an automobile in terms of its role of *transporting passengers*, for example, then we will also need to explain the contexts within which there are passengers waiting to be transported; there can be no transporting role in the absence of passengers[5]. This kind of explanation counts as interactive since it includes the larger environment within which automobiles function and seeks to explain its relational features. On the other hand, if we zero in on what the components of a motor vehicle are, asking how each functions, what its individual features are, and so on, we will be providing a homuncular explanation. Each of these explanations will deepen our understanding of the car concept because each will explain a different aspect of it; such explanations are, thus, not incompatible.

But we have to be careful; it's easy to apply the wrong type of explanation to a concept, as we do when we give a homuncular explanation to a collective concept and vice versa. To see how quickly this confusion can occur, let's look more closely at the homuncular description of *automobile*. Being homuncular, it will focus on car components and on how the various mechanisms function to bring about system-level activity such as acceleration, deceleration, and so on. But note that a concept like *acceleration* is a collective concept, since it applies only to

---

[4] Craver and Bechtel (2007) also do an excellent job of clarifying some of the level confusions that abound in the debate between bottom-up and top-down causation. Much of what they say is mirrored in what Clark says and what I am arguing for here, but to make those connections explicit would take more space than I have room for so I leave that to another paper.

[5] Of course, absence here cannot mean that there just don't happen to be passengers here at this time; rather, it means that the kind *passenger* just doesn't exist in this context.

the car as a whole and involves relational attributes such as the property of increase in speed relative to a frame of reference. Thus, although we could pick out the engine as a mechanism that plays a role in the car's capacity for acceleration, we will need to be careful that we don't erroneously, or sloppily, treat the engine as the car's *accelerator.* The engine itself doesn't do any accelerating at all, it doesn't even move, even though its activity, in conjunction with the movement of the wheels, the amount of friction between the tires and the road, and so on, results in the car's acceleration. In other words, its actions are necessary for acceleration, but the engine itself does not accelerate. The mistake we make, if we take the engine to *be* the car's accelerator, is to give a homuncular explanation to a collective concept: no amount of component activity could ever give account of the relational attributes of such a concept. Thus, homuncular explanations are good for explicating the individual (aspects of) concepts and interactive explanations are required for providing an account of the collective (aspects of) concepts.

We now have the terminology we need to clearly identify the problem with treating neurons as detectors. The concept of being a detector is a collective concept and, as such, requires explication in interactive terms. When we appeal to the detecting capacity of our neurons in the course of explaining the representational capacities of human cognitive agents, however, we are giving a homuncular account, since neurons are components of this larger system. But, since the concept we are trying to explicate is a collective one, this can't possibly be right. Supposing that our capacity to model objects in our environment is explained by the capacity of our neurons to do exactly that is like pointing to a car's engine and saying "there, that's where the acceleration is happening." As we just saw, although car engines play a role in acceleration, to fully explicate the concept we need to look beyond the car's components to the general environment within which acceleration becomes possible. In a precisely analogous way, we shouldn't look inwards for detectors; we need to think more broadly about what contextual attributes make the role of detection possible.

But, someone might counter, why couldn't the neuronal level really be the locus of detection in the human cognitive system? If we suppose that there is a larger system, perhaps a network of neurons, within which neurons function as detectors, we are giving an interactive explanation of the capacity.

Unfortunately, this won't work: if we suppose that neurons function to carry information about whatever it is they detect within the context of a larger system, then we will need to explain how this larger system has the capacity for *using the information* the neurons carry. Otherwise, we'll be back to square one, as we were at the end of section one. But to suppose that something is capable of using information is just another way of saying that it has intentional capacities, that it has the ability to extract a representation of an actual or possible state of affairs from some causal regularity. Such an account would be viciously circular since the very reason we are appealing to the supposed detection capacities of neurons is to explain how the larger system, the human cognitive agent, manages to represent.

Stated thus, this result might seem hopelessly depressing, but I think it is cause for optimism: clearly seeing the circularity of our current thinking ought to liberate us once and for all from whatever reductive attractions it holds. In the next section I will sketch what I see is the way forward.

Section III: A Paradigm Shift

Ramsey's arguments uncover some very deep-seated assumptions about representation that we, perhaps because we are paradigm examples of information-using systems, all seem to share. These biases lead us to read more into causal relations than are justified – co-relation between two states of affairs is not enough to warrant the assumption that information transfer plays a role in the underlying causal transaction, *even* when there is a story to tell about how having and using the relevant information *would have* bestowed selectional advantage on the system within which such states exist.

When we analyse our theoretical approach further, we find that it is underwritten by a confused understanding of the relation between concepts and explanations, that our (natural) reductive impulse to prefer homuncular explanations draws us to look inward when we are explaining intentional capacities when we should be looking outward for interactive explanations instead.

Interaction theorists, and dynamic systems theorists in general, have begun developing precisely these kinds of interactive explanations. (Freeman, 2000; Keijzer, 1998; Kirsch, 1990; Thelen, Schöner, Scheier, Smith, 2001). Among these, Fred Keijzer's is particularly noteworthy since he has attempted to give at least the beginning of an account of the kind of higher-level, off-line behaviour – planning, remembering, and so on – that interactionist accounts with their emphasis on system-environment interactions, have had a hard time explaining. What's particularly exciting about his idea is that it draws its inspiration from the field of genetics, an area in which a paradigm-shift away from representation-based models is already yielding fruitful new insights. On this new view,

> Genes do not instruct the cytoplasm, they rely on the intrinsic disposition of cytoplasmic processes to generate spatial and temporal structure. As Gottlieb puts it, genes are a part of a complex but highly coordinated system of regulatory dynamics that operate simultaneously at multiple scales, extending from genes to chromosomes, to the cell's nucleus, cytoplasm, tissues and up to the whole organism (Gottlieb, 1992, p.142). (Keijzer, 1998, pp. 286-87)

If we are to progress in our understanding of cognition, Keijzer argues, we need to similarly replace our homuncular treatment of behaviour as ultimately driven by internal representations, implemented by neurons, by an interactive theory of how the different scalar levels of activity within and without a cognitive agent influence and direct one another to produce behaviour. On such an account, we are free to understand the function of neurons in entirely novel ways. Keijzer describes one possibility like this:

> In behavioral explanations based on representational specification the activity of neurons is interpreted as an input-output device which receives and sends information. However, neurons can also easily be interpreted as oscillatory units (Alexander & Globus, 1996). Given this interpretation, the total nervous system forms a larger oscillatory network, the behavior of which depends on the characteristics of its components and their connections. As the nervous system is an organ that extends itself over the scale of the total body of an organism, and because the connections between neurons allow very swift interactions across this network, it forms a means for dynamical patterns to organize themselves very fast (starting at tens of milliseconds) at the bodily scale. In turn, the neural dynamics is tied to a musculo-skeletal system capable of initiating environmental changes at the bodily scale. The bodily dynamics in turn influences dynamical relations within the environment. (1998, p. 279)

Whether or not this is ultimately the right way of thinking about neurons is beside the point of this paper; I present it here simply as an example of the theoretical possibilities open to us.

I'll leave the final word to Walter Freeman, a neuroscientist who claims that he was able to make headway in interpreting his own data only once he let go of his basic assumption that neurons function as detectors:

> For more than 10 years we tried to say that each spatial pattern was like a snapshot, that each burst served to represent the odorant with which we correlated it, and that the pattern was like a search image that served to symbolize the presence or absence of the odorant that the system was looking for. But such interpretations were misleading. They encouraged us to view neural activity as a function of the features and causal impact of stimuli on the organism and to look for a reflection of the environment within by correlating features of the stimuli with neural activity. This was a mistake. After years of sifting through our data, we identified the problem: it was the concept of representation. (Freeman & Skarda, 1990, p.376)

References

Bowers, Jeffrey S. (2009). On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychological Review, Vol. 116(1)*, 220-51.

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

Clark, A. (1996). Happy couplings: emergence and. explanatory interlock. In M. Boden (Ed.), *The Philosophy of Artificial Life*. Oxford: Oxford University Press.

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

Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, *22*: 547–563.

Dehaene, S. (2002). Single-Neuron Arithmetic. *Science*, *New Series*, *Vol. 297*, *No. 5587*, 1652-1653.

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

Dretske, F. (1994). The Explanatory Role of Information, *Phil. Trans. R. Soc. Lond. A*, 349, 59-70.

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

Freeman, W., & Skarda, C. (1990). Representations: who needs them? In J. McGaugh, N. Weinberger and G. Lynch (Eds.), *Brain Organization and Memory: Cells, Systems and Circuits*. Oxford: Oxford University Press, pp. 375–380.

Freeman, W. (2000). *How Brains Make Up Their Minds*. New York: Columbia University Press.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., Tenenbaum, J. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences 14*, 357–364.

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

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

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

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

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

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

Thelen, E., Schöner, G., Scheier, C., Smith, L. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, 24, 1–86.