

# Mechanistic and ecological explanations in agent-based models of cognition

Jason Noble<sup>1</sup> and Manuel de Pinedo<sup>2</sup>

<sup>1</sup>School of Computing, University of Leeds, UK

<sup>2</sup>Department of Philosophy, University of Granada, Spain

jasonn@comp.leeds.ac.uk

## Abstract

We argue that two styles of explanation—mechanistic and ecological—are needed in accounting for the behaviour of synthetic agents. An emphasis on mechanistic explanation in some current ALife models is identified, and parallels are drawn with issues in the philosophy of mind literature. We conclude that ecological or agent-level explanation does not come with representational baggage, and that mechanistic explanations cannot stand alone.

## Explaining the behaviour of artificial agents

This paper is concerned with the ways in which artificial life researchers explain the behaviour of their evolved agents. We see a role for both mechanistic explanations, and ecological or agent-level explanations. In a mechanistic explanation, the agent is accounted for by showing how its component parts interact with each other and the environment to result in the production of the behaviour in question. In an ecological or agent-level explanation, the mechanical details are abstracted away and a behaviour is explained by saying what its purpose or function is *for the agent*—there may be explicit or implicit reference to a history of learning or selection.

We are interested in exploring the use of these two types of explanation for two reasons. First, we see fruitful parallels between artificial life and philosophy of mind in this respect. Behaviour and its explanation are central to both disciplines, and we hope to show that each area has something to learn from the other. Second, we are worried about a tendency for mechanistic explanations in artificial life to be regarded as somehow more basic or more primary. This tendency mirrors the rise in philosophy of mind of an idea known as *eliminativism*, which seeks to replace agent-level talk about behaviour with a mechanistic vocabulary. The eliminativist holds that explaining a person's behaviour at the agent level (e.g., explaining John's trip to the supermarket by saying that he wanted to buy milk) is pre-scientific, and that such accounts will one day be replaced by complex but precise neurological explanations. We believe that eliminativism is a deeply flawed view. We will argue that

artificial life research should not only avoid the temptations of eliminativism, but can actually help in showing what is wrong with the idea.

Artificial life research covers a broad territory, but it can reasonably be summarized as seeking to understand life and mind through synthetic means. We want to focus on a particular strand of artificial life work in which mechanistic explanations are sought for the behaviour of evolved agents. In a typical paper of the type we have in mind, an evolutionary algorithm is applied to simulated or robotic agents with a flexible control architecture (often a recurrent neural network). The fitness function measures performance on some task, and eventually the agents achieve high levels of performance, perhaps using an unanticipated strategy. An elite or typical agent is then analyzed in detail and a mechanistic explanation of the behaviour is presented (e.g., an elaborate description of just how the weights and connections in the neural network facilitate the successful behaviour). The authors of the paper typically conclude that the behaviour is now fully explained: the artificial agent's capacity to perform behaviour X in environment Y is accounted for by having a control architecture tuned just so. The inference is often drawn that naturally occurring agents (animals) performing a similar behaviour might use a similar mechanism.

This tradition in artificial life research can be traced back to the seminal work of Walter (1950), and was revived by Braitenberg (1984), who demonstrated that behaviour seemingly requiring complex agent-level explanations could be produced by some very simple mechanisms. (Braitenberg was playfully agnostic, however, on the question of whether his models meant that agent-level explanation could be dispensed with entirely.) Braitenberg's work has been extremely influential in artificial life. Research that has followed on in a similar vein includes that of Beer (1990, 1996) who is involved in a long-term program to produce models of "minimally cognitive behavior" and to analyze the evolved architectures using a dynamical systems approach. Other work that fits our template includes Cliff, Harvey, and Husbands (1993), Floreano and Mondada (1994), Husbands, Harvey, and Cliff (1995), Quinn, Smith, Mayley, and Hus-

Copyrighted Material

bands (2002), and others too numerous to mention. Webb's (1994) paper is an interesting example in that the evolved agents to be analyzed are not synthetic but real insects; the work uses a robotic analogue of the cricket to demonstrate the plausibility of a simple mechanical explanation for sonotactic behaviour.

The work of Randall Beer epitomizes the approach we are discussing, and so from here on we will focus on his most recent paper (Beer, 2003) which presents a highly detailed dynamical systems analysis of the workings of an evolved agent. The agent is controlled by a continuous-time recurrent neural network, it is equipped with a fan-like array of simple range sensors, and it can move along a one-dimensional track while two kinds of objects fall from the sky above. The agent is the elite member of a population that has been selected for the ability to catch falling circles and to avoid falling diamonds. Beer's analysis demonstrates that the catching and avoiding behaviour can be understood as the result of a complex dynamical system that includes the agent's neural control architecture, its simulated body, and its environment. Beer presents this type of analysis as a potential new paradigm for explanation in cognitive science.

We should reassure the reader at this point that we are admirers of Beer's work and of that of the other authors cited above. Mechanistic explanations of artificially evolved agents (or of naturally occurring agents) are valuable. Our problem is with the implication that mechanistic explanations are independent and exhaustive explanations, containing all the conceptual resources necessary for understanding behaviour. We will expand on this point below, but it seems to us that such explanations are only interesting inasmuch as they shed light on broader questions about life, cognition, and the way in which an agent is situated in its environment.

One of the reasons for artificial life research to have focussed on this sort of mechanistic explanation is because of the way in which the field has defined itself in opposition to the classical artificial intelligence tradition ("GOFAT"). In classical AI, an agent's behaviour is explained as the result of it planning a course of action based on its internal model of the world. The agent is supposed to be using sensory input to update its representation of what is going on in the world, and then to be manipulating these internal representations in order to plan and re-plan the optimal way to achieve its explicit goals. ALife evangelists such as Brooks (1991) and Harvey (1996) were quick to point out problems with this picture, and to supply counter-examples in the form of simple reactive agents that were capable of highly competent behaviour without possessing anything resembling an internal representation. If you could completely describe the workings of an agent's neural network in simple mechanistic terms, as did Cliff et al. (1993) for example, you could then ask, "Where are the internal representations?" Nowhere, of course.

Beer's (2003) paper is also driven by this ongoing objec-

tion of representationalism: Beer argues that dynamical systems explanations (i.e., mechanistic explanations) raise "important questions about the very necessity of notions of representation and computation in cognitive theorizing" (p. 3). But the rejection of representationalism leads Beer to take for granted a false dilemma. He associates explanations cast at the whole-agent level—such as an explanation of the agent's movement in terms of a belief that the falling object was a diamond—with a representational perspective. In other words, if the agent can be sensibly described as believing that a diamond is falling, then there had better be a diamond-detecting routine and an internal switch or memory register to represent the detected diamond. Alternatively, it might be possible to explain the agent's behaviour in terms of mechanistic properties of the coupled brain-body-environment system. This is the false dilemma. Obviously Beer makes a strong case for the second option: after close examination of the circle-catching, diamond-avoiding agent, we find no circle or diamond detectors and nothing that resembles a representation of a circle or a diamond. Beer concludes that internal representations are not needed to explain the agent's behaviour, whereas the right set of dynamical equations allows us to understand what is going on.

Representationalism takes something of a battering in Beer's paper. We end up with a bold new vision for the explanation of behaviour in terms of dynamical systems theory, with none of the familiar agent-level explanatory concepts—beliefs, desires, intentions—in sight. This is clearly at odds with our own view that both kinds of explanation are desirable and necessary. We will therefore question Beer's assumption that agent-level explanations and internal representations go hand in hand. In order to do so, we must first look at why representationalist views held any appeal in the first place.

Representationalism is just one aspect of a bigger and older idea, namely *internalism*: roughly speaking, this is the belief that cognition is something that goes on inside the head. And as so often happens in issues to do with the mind, the problem is all down to Descartes. Internalism is attractive because it matches the Cartesian intuition that thoughts, cognition, and knowledge are processes or things located in a special, non-physical place called the mind. Putting it another way, representations are theoretical devices to give flesh to the intuitive appeal of the idea that thought happens inside the agent—in its head, soul, brain, or wherever. The initial mistake is to suppose that a lonely homunculus inhabits the mind, and the mistake is compounded by bringing in representations for the homunculus to calculate over. Thus intuitive internalism gives rise to representationalism and not the other way around. It follows that representationalism is best dealt with at its source: that is, by challenging the notion that thought is internal.

There is some irony here. In an important sense, artificial life is the last place to expect internalism. Unlike artificial

intelligence, which smacks of internalism in its resolve to keep mind and world separate, artificial life researchers have been quick to appreciate that cognition can be distributed across the agent-environment divide. Witness the popularity of the key phrases “situatedness” and “embeddedness” in the ALife literature. So we need to be clear about what we are claiming: that the diagnosis of representationalism’s flaws has not been radical enough. A last vestige of internalist thinking has resulted in an incorrect association between representationalism and agent-level explanations, and this in turn has produced an unwarranted emphasis on low-level mechanistic explanations of artificial agents.

### Parallels in philosophy of mind

We want to pause momentarily in our analysis of Beer (2003) and move to the philosophy of mind literature. Philosophers of mind are much concerned with the explanation of behaviour, and their nearest equivalent to the agent-level / mechanistic distinction is that between personal and sub-personal explanations. This was informally introduced by Ryle (1949) and later popularized by his student Dennett (1969). In *The Concept of Mind* (1949) Ryle argued that there are two very different classes of human behaviour requiring explanation. When we get something right, our behaviour is best explained at the personal level, in the ordinary language of beliefs and desires: John drove to the supermarket because he desired milk and believed that it was available there. On the other hand, when we get something wrong, a sub-personal explanation, phrased in terms of physical interactions between our component parts, may be called for. Suppose John crashes his car on the way to the supermarket, because he suffers a mild stroke. The relevant explanation is obviously sub-personal.

An important point to appreciate about the two kinds of explanation is that personal-level stories stop quite early. Once we have granted that John is a rational agent, there is not much more to say about his trip to the supermarket. Ryle was content with this short chain of reasons for rational acts, and was extremely sceptical about the idea that psychology or any other science could somehow supplant personal-level explanations and supply the ‘real’ causal story behind a person’s actions. Sub-personal explanation, on the other hand, can go very deep: accounting for John’s stroke might involve considerations of diet, physiology, genetics, biochemistry, and ultimately physics. Sub-personal explanation is also part of the story in explaining John’s competencies: how is it that he has the sensorimotor coordination needed to drive a car, or to remember the way to the supermarket? It was in these sorts of questions that Ryle saw a role for cognitive science.

Furthermore, Ryle held that the category errors generated by confusion between the two different levels of explanation were responsible for most of the apparent mysteries about cognition. Ryle’s most famous example of a category error involves a visitor being shown all of the buildings in Ox-

ford and then insisting “Yes, but where is the university?” A similar error is easily committed when thinking about agents and their component parts. For example, John is like the university: he is in a sense constituted by his component parts, but he is not the same type of thing as one of his components. Forgetting this, and imagining that the components of John’s brain and John the person are on a par with each other, leads to conceptual disasters such as the ‘mystery’ of how John could possibly be conscious, for example. (If John is equated with the mere matter of his brain then certainly there is a mystery about how the-matter-that-is-John achieves conscious awareness, but if we recognize that John and his brain matter are concepts at different levels, then wondering about how John could possibly be conscious is properly recognized as being a bit like asking how it is that a university can make decisions despite being constructed from stone.)

Regarding the issue of internalism, it is worth noting that Ryle also embraced a radical *externalism*. Ryle argued that cognition is spread out across the agent and the world, and wanted to banish the Cartesian image of mind as a private place (whether non-physical or physical) where thoughts happen. A similar externalist view has been adopted by many more recent authors in the philosophy of mind (see, e.g., Morris, 1992; McDowell, 1998) and seems in obvious harmony with the concern for situatedness and embodiment in artificial life.

Perhaps because of his accessible style and engagement with the sciences, Dennett has been an influential philosopher in the artificial life community and so a further word on his position is in order. Dennett (1969) first espouses a view very similar to Ryle’s, but in later work (Dennett, 1987) he retreated somewhat and claimed that personal-level or agent-level explanation is just a stance that can be taken when it is expedient to do so. However, Morris (1992) points out that this is plainly self-defeating, as the idea that there is anyone around to take stances toward anything presupposes the existence of persons or agents. Our own view is that Ryle’s original distinction is worth defending.

We do not wish to give the impression that the personal / sub-personal distinction is universally accepted in the literature. Many authors have opposed it; the argument usually goes as follows. Even if personal-level explanations are taken to be somehow autonomous, as in Ryle’s account, then more or less often we will still need sub-personal explanations to fill in the gaps, i.e., to explain what went wrong when rational action fails. If we can be successful in explaining *some* behaviour at the sub-personal level, then surely there is a prospect of explaining all behaviour this way. It would therefore be parsimonious to eliminate reference to persons (as well as to beliefs, desires, reasons, etc.) altogether. Behaviour is then to be explained by something like neuroscience. This is of course the eliminativist view: see Churchland (1981) for the classic account. In the next

section we will explain why we believe eliminativism to be untenable.

Philosophers of mind have tended to focus on the minds and behaviour of normal adult humans. As a result, the status of actors occupying the broad space between thermostats and rational adults (e.g., animals, infants, people with neurological disorders, etc.) can be problematic for otherwise excellent accounts of agency. This seems to be an area where philosophy of mind could look to artificial life for richer ways of categorizing the various possible classes of agent (see Dennett, 1996, for an example of this). And although the personal / sub-personal distinction is a useful one, McDowell (1998) points out that its use in conjunction with a focus on adult humans has led to a confusion between what should be two separate distinctions. Personal-level explanation seems to belong to a realm of rational, normative agents, whereas sub-personal explanation concerns the disenchanted physical world. McDowell draws on the landmark neuroscience paper “What the frog’s eye tells the frog’s brain” (Lettvin, Maturana, McCulloch, & Pitts, 1959) and suggests that to fully understand the frog, we would also need a “froggy / sub-froggy” distinction. On the one hand we need to consider the frog as a whole agent in its environment or *Umwelt*, and look for example at the significance of different environmental features *for the frog*. This corresponds to what we have called agent-level or ecological explanation, and McDowell aptly cites the ecological psychology of Gibson (1979) to illustrate the approach. On the other hand we will also want to employ familiar styles of mechanistic explanation in looking at how various components of the frog (the visual system and the motor system for example) interact with each other. McDowell references Marr’s (1982) work on computational models of vision as an example of the latter approach.

In none of this does McDowell (1998) argue that frogs are rational, normative agents. McDowell’s conclusion is that there are really two distinctions at work: one between agents and their component parts, and one between persons and non-rational agents. We endorse McDowell’s argument, and stress that we are interested here in the first of the two distinctions. We see the split between personal and sub-personal explanation in philosophy of mind as a special case of the more general distinction between agent-level or ecological explanation, and mechanistic explanation. What of the second distinction? It certainly points to a profound philosophical question about what it might be that distinguishes a person from a preying mantis. Philosophers such as Morris (1992) and many others have attempted to answer that question in terms of qualities such as moral responsibility, but we plan to remain silent on the issue. It seems to be a problem about which artificial life does not yet have anything sensible to say.

## Explanatory pluralism needed in artificial life

We believe that artificial life requires both mechanistic and ecological explanations in order to make sense of the behaviour of synthetic agents. Our argument, in brief, is that agent-level explanations are necessary because without them mechanistic explanations are incomprehensible, that any suspicion that agent-level explanations might be reducible to mechanistic explanations is founded on a metaphysical error, and that an externalist perspective means there is no danger of ecological or agent-level talk being linked with representationalism.

We start by returning to our critique of Beer (2003). Beer seems to believe that because he has rejected representationalism, he is left with an eliminativist position with respect to agent-level concepts such as knowledge, meaning, belief, and desire. In other words, there is no place for such concepts in cognitive science, and the only proper explanations will be mechanistic, dynamical systems accounts of agents coupled with their environments. However, this would only follow if the representationalist perspective was the only way to make sense of knowledge and meaning. A sufficiently radical externalist perspective on these concepts does away with representationalism without disposing of all agent-level talk. Consider John and his trip to the supermarket for milk. To say that John intends to drive to the supermarket simply does not mean being committed to the notion that John has an explicit mental representation of a planned route mapped out in his head. Presumably John will get there through some distributed combination of driving habits, consulting a street directory, reactive strategies at particular intersections, environmental features such as street signs prompting the right behaviour, etc. In terms of the explanatory utility of the agent-level description of his action, the one phrased in terms of his desire for milk and his belief that the supermarket sells milk, it does not really matter. It would be interesting to study John’s sub-personal capacities and find out how they constrained the sorts of milk-buying journeys he might be capable of, but such mechanistic explanations will not lead to a revision of the agent-level story: on this occasion, he went to the supermarket to buy milk.

The dilemma identified by Beer (2003) is thus a false one, and externalism means that it is possible to save the baby of agent-level explanation while throwing out the representational bathwater. Artificial life and cognitive science generally do not need to give up on talking about agent-level concepts such as knowledge or meaning: practitioners in both fields just need to recognize that knowledge and meaning, as much as perception and action, are features of the coupled agent-environment system and not something internal.

Why do we feel that agent-level description is of value? Why are we convinced that the impressive analytical vocabulary of dynamical systems theory, for example, is not the only vocabulary needed by the artificial life researcher? We

refer the reader to the deceptively obvious fact that Beer (2003) needs to describe his agent as a circle catcher and a diamond avoider. Indeed, these are the propensities that his agent was selected for over many generations of evolution. This description is admittedly simple, but it is agent-level talk, and clearly of a different explanatory level than a description of the agent / environment system in terms of differential equations. As a quick thought experiment of our own, we ask whether anyone could possibly make sense of the behaviour of the agent given only the dynamical systems description so carefully developed in Beer's paper, and not the brief but enormously helpful agent-level description. Looking only at the mechanistic level, it would be extremely difficult and perhaps impossible to see that all of this complexity was in the service of circle catching and diamond avoidance.

McDowell (1998) makes the same point in slightly different language. At the agent or ecological level we can pose and answer "why?" questions. Why did the frog stick out its tongue? In order to catch what it believed was a fly. These questions and answers can in turn inspire "how?" questions at the mechanistic ("sub-froggy") level. How did the visual input lead to the appropriate motor output? When we have answered the mechanistic how-question in terms of some sort of neural circuitry diagram, we have described what McDowell calls an enabling condition for the agent-level behaviour. If we were to then insist that this mechanistic explanation could stand alone, we are mistaking an enabling condition for a constitutive one. As Davidson (1980, p. 247) puts it "... it is one thing for developments in one field to affect changes in a related field, and another thing for knowledge gained in one area to constitute knowledge of another." Even the best mechanistic explanation will be incomprehensible without an agent-level framework. If systems as simple as the one analyzed by Beer (2003) require on the one hand agent-level explanations and on the other hand a mechanistic description in terms of dynamical systems, then clearly more ambitious targets such as advanced ALife agents, frogs, and human beings will also require both levels of description.

We have said that we would show why eliminativism is wrong. Eliminativism argues that agent-level descriptions are no better than myths, and stand in need of elimination in favour of mechanistic explanations. It is in the kind of practical interplay between agent-level and mechanistic explanations described above that we think artificial life demonstrates why the former cannot be eliminated in favour of the latter. But where did anyone ever get the idea that an ecological or agent-level explanation would be eliminable by a mechanistic one? This is an instance of a widely held and often unquestioned belief that all forms of explanation will eventually be replaced by or reduced to one privileged explanatory basis, usually assumed to be the language of physics. Morris (1992) identifies this belief as *scientism*, and demolishes it in short order. Morris is in no way anti-

scientific, but he disagrees with a movement in philosophy known as *naturalism*, which seeks to use the methods of natural science as a basis for metaphysics, i.e., as a basis for thinking about what exists and about how we could know about it. The first move of the naturalist is to propose the sciences (and ultimately physics) as the only basis for knowledge, and to declare "There are only scientific facts." Morris points out that this move is immediately fatal: what kind of fact is the declaration itself? It is clearly not a testable proposition from the natural sciences, and the declaration thus perversely renders itself false.

We see eliminativism as one of the faces of scientism. The onus of proof is not on the user of agent-level explanations to say why they are autonomous with respect to mechanistic explanations. Rather the burden runs the other way: given the failure of scientism, the eliminativist must show why and how an agent-level explanation could be dispensed with. It is not valid to simply assume that mechanistic explanations are primary.

If our argument holds, then artificial life should be content to deal in multiple levels of explanation for the behaviour of synthetic agents. Undoubtedly some researchers will react to this assertion with horror, whereas others will shrug as they are already committed to such pluralism. By way of reassurance for the first group, we want to point out that there is a strong precedent for the peaceful coexistence of multiple levels of explanation in a scientific discipline: we refer to Tinbergen's (1963) seminal paper on the aims and methods of ethology. Tinbergen introduced four types of explanation for ethology, and arguably for biology in general: two of them were explanation in terms of function and in terms of mechanism, which obviously correspond to the two types of explanation we have been discussing. Tinbergen's two additional types of explanation were both historical: ontogenetic and phylogenetic explanation. The good news is therefore that one of artificial life's parent disciplines appears to be able to cope with a plurality of explanatory projects.

In conclusion, we feel that Beer (2003) should not be concerned that endorsing an agent-level description will commit him to the follies of old-fashioned representational AI. One can say that Beer's agent catches circles and avoids diamonds without conceiving of cognition as a series of rule-governed operations over internal symbols. We encourage Beer (and others in artificial life) to go all the way with the externalism exhibited in his analyses of perception and decision-making. On the view we are urging, mind is not internal, it is all over the place; indeed, "mind" is just a very abstract way of describing the agent / environment interaction. The debate over representationalism loses all urgency once the Cartesian image of the mind as a place of internal knowings has been properly dispelled.

## Anticipated objections

We have anticipated two of the more likely objections to our argument and have attempted to answer them in advance.

Some might see our argument about the need for two kinds of explanation as a pragmatic move related to the difficulty of understanding the messy architectures of evolved agents. In contrast, when dealing with a traditionally engineered system (a hand-coded AI robot for example) perhaps only the mechanistic level of explanation will be necessary, as a complete and accurate blueprint of the agent's architecture is available. We disagree: such an agent will still need to be understood in terms of the mechanical interactions between its components, and at the agent-level in terms of the designer's intentions. Random wandering in a vacuuming robot, for instance, might be intended to clean the carpet, independently of the way in which that movement is implemented at a lower level of description.

Accusing your opponents of the hangovers of Cartesian thinking is a popular sport in cognitive science and philosophy, and we too might be faced with the accusation that in proposing two levels of explanation for agents we are resurrecting some sort of dualism between a physical, mechanistic, sub-agent domain and a mysterious, cognitive agent-level domain. The important thing to emphasize here is that we are not positing two kinds of thing, physical-stuff and mind-stuff. The insistence that to each variety of explanation corresponds a variety of stuff is itself a Cartesian idea. Our externalist, Rylean, perspective precludes seeing the mind as a place or as a special sort of thing. For us, mind is shorthand for a set of complex interactions in a unified world that happen to demand their own sort of explanation.

## References

- Beer, R. D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., & Wilson, S. W. (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pp. 421–429. MIT Press / Bradford Books, Cambridge, MA.
- Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11(4), 209–244.
- Beer, R. D. (1990). *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology*. Academic Press, Boston.
- Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78(2), 67–90.
- Cliff, D., Harvey, I., & Husbands, P. (1993). Explorations in evolutionary robotics. *Adaptive Behavior*, 2(4), 371–380.
- Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.). (1994). *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*. MIT Press / Bradford Books, Cambridge, MA.
- Davidson, D. (1980). The material mind. In *Essays on Actions and Events*, pp. 245–259. Clarendon Press, Oxford.
- Dennett, D. C. (1969). *Content and Consciousness*. Routledge & Kegan Paul, London.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press / Bradford Books, Cambridge, MA.
- Dennett, D. C. (1996). *Kinds of Minds: Towards an Understanding of Consciousness*. Weidenfeld & Nicolson, London.
- Floreano, D., & Mondada, F. (1994). Automatic creation of an autonomous agent: Genetic evolution of a neural-network driven robot. In Cliff et al. (1994), pp. 421–430.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.
- Harvey, I. (1996). Untimed and misrepresented: Connectionism and the computer metaphor. *AISB Quarterly*, 96, 20–27.
- Husbands, P., Harvey, I., & Cliff, D. (1995). Circle in the round: State space attractors for evolved sighted robots. *Robotics and Autonomous Systems*, 15, 83–106.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers*, 47, 1940–1955.
- Marr, D. (1982). *Vision*. W. H. Freeman & Co., San Francisco.
- McDowell, J. (1998). The content of perceptual experience. In *Mind, Value and Reality*, pp. 341–358. Harvard University Press, Cambridge, MA.
- Morris, M. (1992). *The Good and the True*. Clarendon Press, Oxford.
- Quinn, M., Smith, L., Mayley, G., & Husbands, P. (2002). Evolving teamwork and role allocation with real robots. In Standish, R. K., Bedau, M. A., & Abbass, H. A. (Eds.), *Artificial Life VIII: Proceedings of the Eighth International Conference on Artificial Life*, pp. 302–311. MIT Press, Cambridge, MA.
- Ryle, G. (1949). *The Concept of Mind*. Hutchinson, London.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410–433.
- Walter, W. G. (1950). An imitation of life. *Scientific American*, May, 42–45.
- Webb, B. (1994). Robotic experiments in cricket phonotaxis. In Cliff et al. (1994), pp. 45–54.