440

Natural Kinds, Autonomous Robots and History of Use

Horst Hendriks-Jansen
School of Cognitive and Computing Sciences
University of Sussex
email: horsth@cogs.susx.ac.uk

Abstract

Since the publication of Dennett's *Content and Consciousness* (Dennett, 1969), cognitive science has relied, sometimes explicitly and often implicitly, on Darwin's Theory of Natural Selection to 'naturalise' the explanatory entities it borrowed from cybernetics and A.I. The assumption has been that internal representations, derived top-down through functional analysis of intentional behaviour, can be given an evolutionary pedigree and rationale, thereby elevating them to legitimate natural kinds for psychology. Millikan's (1984) appeal to 'history of use' represents the most comprehensive and radical attempt to provide a philosophical basis for this assumption. I argue in this paper that, although Millikan's strategy is correct, it cannot be applied to internal representations. I suggest that it may be used instead to identify and legitimise certain crucial patterns of situated activity, which are not reducible either to features in the world or to neurophysiological events, and that these should be considered the natural kinds of psychology. The argument is based on a detailed examination of an autonomous, navigating robot designed by Mataric (Mataric & Brooks, 1990).

1 The Need for Natural Kinds in Psychology

It is increasingly being argued that cognitive science, conceived as a legitimisation of folk-psychology and the 'intentional stance' (Dennett, 1978) through computer-based models, has failed to provide a satisfactory scientific explanation of human and animal behaviour. My interest lies in a particular version of this argument. It centres on the fact that A.I. models, both of the 'classical' and the connectionist type, depend on internal representations which are derived top-down from an analysis of full-blown symbolic behaviour, and that such entities do not constitute a defensible *natural kind.*

Natural kinds may be thought of as the scaffolding for scientific theory building. Hypothetical groupings or explanatory entities, suggested by a growing theory, lead to fruitful inductions and so become established as part of a new scientific paradigm.

It has been suggested (e.g. by Wilkes, 1989) that one reason why the behavioural sciences are underdeveloped is their inability to arrive at a concensus about natural kinds. No acceptable, shared taxonomy exists in psychology. It does not have a paradigm in Kuhn's (1970) sense. Cognitive science can be seen as a major attempt to construct such a paradigm, proposing internal representations as a natural kind of psychology.

The notion of natural kinds, and a recognition of their importance in scientific advance, are quite compatible with scepticism as to whether such hypothetical groupings are in any sense 'real' constituents of the world. 'Naturalness' may be glossed in terms of the likelihood that the chosen kinds can generate fruitful laws and generalisations. However, the expectation must be that those laws and generalisations will in some way succeed in tying the hypothetical explanatory entities into the rest of science.

In **Content and Consciousness**, Dennett (1969) suggested that the Theory of Natural Selection could be made to perform that function for concepts derived from A.I. and cybernetics. The challenge of showing how this might be done has been taken up by, among others, Dretske (1988, 1992), Sterelny (1990) and Millikan (1984, 1989). I believe that none of these authors has finally succeeded in bridging the gap - that internal representations cannot be 'naturalised' by an appeal to natural selection, since any attempt to do so must resort to an overly simplified view of the genetic and environmental factors involved and will require a realist ontology in terms of objects, properties and situations which cannot be sustained (see Churchland, 1989 for a detailed discussion of this problem).

It might be said that cognitive science represents a coming together of two ideas which have made it impossible to discover a helpful explanatory entity for the phenomena of mind. One strand holds that we are the product of evolution; the other that the mind is in some sense like a computer, and ought therefore to be conceptualised as a general-purpose machine which runs particular bits of software to perform specific tasks. There is nothing intrinsic to this machine, nothing inherent in the explanation of its mechanism, which relates it to the world. If its congruence with the world is to be explained, additional explanatory entities, in the form of internal representations, need to be introduced. The question is how these can be legitimised, tied in to the rest of science: 'naturalised' or 'grounded'.

This is the conundrum faced in various ways by Dennett (1969), Dretske 1988, 1992), Sterelny (1990), Millikan (1984), Harnad (1989), Smolensky (1988) and others. But the puzzle is an artificial one. It stems, I would like to suggest, from an imposition of the wrong natural kinds, which in turn appear to be required because the brain is conceived as a stand-alone, all-purpose mechanism whose essential operations may be specified without reference to the circumstances under which they evolved. This all-purpose mechanism encourages formal task-description. Anything which can be formalised as a task-description or effective procedure may be run on a computer. But neither that fact, nor the logic of the program that is written to implement it, nor the type of virtual machine on which it is run, confers compelling explanatory force. As Heil (1981) points out, we may grant that what a person does or learns is *describable* in a certain way without assuming that the person has, in some sense, internalised that description.

Although cognitive philosophy ultimately depends on natural selection to link computational concepts to the rest of science, there is little discussion of how such a link might be defended on biological grounds. References to evolution tend to be notional and abstract, treating the process as a kind of heuristic, and amounting to the claim that mental structures, as well as external form, may be 'designed' by a selective process. The fact that such a process can only operate indirectly on neural structure through its behavioural consequences, and the questions this raises about how those consequences should be conceptualised and related back to neural structure are rarely mentioned.

Any appeal to Darwin's theory must be tempered by a clear understanding that evolution is not a rational designer. Ethologists have shown that there *are* ways of linking *patterns of behaviour* to the environment through natural selection, but this requires careful observation of an animal *in situ*, imaginative reconstruction or isolation in related species of alternatives to the behaviour pattern being studied, and a strong sense of the limitations of the notion of optimality.

'Optimal' solutions in evolutionary terms might be difficult to recognise as efficient solutions in design terms. Natural selection often hijacks organs or patterns of behaviour which developed for one 'purpose' to serve some other, quite different purpose. Traits may be selected because they are genetically linked to other, completely unrelated but reproductively advantageous traits, even though they themselves do not confer any selective advantage. The conditions under which a particular trait was selected in the past may no longer pertain in its present environmental niche. An understanding of evolutionary 'solutions' requires thinking in historical terms, involving interactions between the logic of genetic reproduction and the pressures of an ever-changing environment.

The chances of discovering underlying mechanisms which are the result of evolution through top-down analysis and formal task-description are thus extremely small. The belief that such a strategy can work is founded on pre- Darwinian analogies with machines and vestiges of a teleological view of nature, which conceive a functioning organism as constructed out of parts, each designed for a specific purpose and contributing thereby to the functioning of the organism as a whole. Ruth Millikan's definition of 'proper function' (Millikan, 1984) is the most radical attempt to imbue this pre-Darwinian notion of function with the scientific legitimacy conferred by Darwin's theory. Unlike some other philosophers, Millikan is fully aware of the problems this involves. If function is to be grounded in natural selection, then the function of an organ or trait must be shown to be totally independent of an observer's interpretation. The organ or trait must have a particular function not because we interpret it that way, but for objective reasons that can be derived from natural selection alone. Millikan's solution to this problem is to appeal to a trait's functional history. The 'proper function' of a wild seed, Millikan says, is to develop into a particular species of plant. Statistically this may not happen very often: the 'normal conditions' for it to happen may be satisfied only rarely. Nevertheless, it only makes sense to think of its proper function as developing into a plant, since that is why it evolved into the seed we see today. Its proper function, and the normal conditions under which this operates, define its 'meaning'. Similarly, the tail-splash of the beaver 'means' danger, because only when it corresponds to danger does the instinctive reaction of the other beavers to take cover serve any purpose. That the splash corresponds to danger is a normal condition for proper functioning of the beavers' instinctive reaction to the splash.

I do not believe this strategy can work for internal representations. Even if the 'intentional icon' we identify today through top-down analysis of the creature's behaviour could reasonably be thought of as a hereditary trait, it would require for its true explanation - for the determination of its real 'meaning' in Millikan's terms - a comprehensive account of its entire history of development and modification. Evolution is an undirected, opportunistic process. If history of use is taken seriously (as I believe it should), then all the false starts and changes of use that went into producing the current intentional icon must be taken into account in defining its current 'meaning'. Strictly speaking, by Millikan's criteria, we are not even allowed to appeal to the proper function of a trait displayed by the present generation of animals until it has been passed on to their descendants - for only then will its proper function in this generation have been established.

The question then becomes whether Millikan's strategy can be made to work for some alternative natural kind that might be proposed for psychology. How can it ever be established, by looking at a contemporary functioning creature, what its key functional entities are? How do we get around the problem of proving that its ecological niche has been around long enough to constitute the determining factor? How do we prove that the history of successive adaptations has not imposed its own logic, and that the quirks of hitch-hiking genes and genetic drift have not skewed the results beyond any hope of functional decomposition? How do we know that the 'problem' posed by the environment has been correctly diagnosed? What we are looking for is some sort of stamp of approval which will legitimise one natural kind rather than another, and as Millikan makes clear, this stamp of approval can only be conferred retrospectively.

In the remainder of this paper, I hope to show, by a detailed examination of an autonomous, navigating robot constructed by Mataric (Mataric and Brooks, 1991, Mataric, 1991a, 1991b) that it is possible to make such a claim for certain patterns of situated activity on the grounds that Millikan suggested. Mataric's robot is not the only implementation in which the structured activity provided by wall-following is used as a basis for navigation. See e.g. Nehmzow & Smithers, 1991, 1992, and Pfeifer & Verschure, 1992 for radically different and highly instructive alternatives. I have chosen Mataric's robot as my illustration because it provides a particularly clear example of the philosophical point I shall be trying to make.

2 An Autonomous Agent which Builds Maps by Finding its Way in the World

The design of Mataric's navigating robot is based on Brooks's subsumption architecture (Brooks, 1985, 1991a, 1991b, 1992). It has three distinct layers, each of which operates independently from the others and receives its own data from the robot's sensors. The lowest layer produces the robot's basic form of structured activity, which might be called 'wall-following' or 'boundary-tracing', the middle layer performs landmark detection, and the top layer does the map learning and navigation.

Mataric's implementation of wall-following uses four independently operating reflexes (each of them closely coupled with the robot's environment), called STROLL, AVOID, CORRECT and ALIGN. The main sensors used in these reflexes are sonar range detectors. Her robot has a total of twelve of these detectors, dividing the surrounding 360 degree space into 30 degree wedges. The sonars have low accuracy and a low refresh rate, and these constraints, together with the characteristics of the robot's environment, determined Mataric's implementation of the robot's boundary tracing behaviour. Basically, STROLL prevents head-on collisions when obstacles materialise close-to in the wedges straight ahead of the robot, AVOID implements a side-ways turn if the obstacle is farther away, ALIGN ensures that the robot will not drift far from a lateral surface by comparing side and rear wedges and making compensatory turns, and CORRECT prevents it from losing track of a wall which suddenly veers away from it by comparing the two lateral wedges.

The wall-following behaviour which results is not a pre-determined series of movements; it is an emergent activity with a structure that lies at a level above that of the reflexes which produce it. As is shown by traces of the robot's path, it follows a particular wall and traverses a particular corridor in different ways on successive occasions. Its route will depend on its approach, on noise in its sensor readings, and on the numerous unspecified contingencies of a dynamic environment. But it can be relied upon to follow the wall each time.

There are no explicit instructions within the robot which tell it to follow walls. The notion of a wall does not have an internal representation inside the robot; no formal description or definition of walls is necessary to produce the behaviour. What Mataric has done is to construct an engineering description of walls in terms of the robot's characteristics and limitations, and to implement that description as STROLL, AVOID, ALIGN and CORRECT in combination with the robot's sensori-motor equipment. The only place at which the underlying mechanisms that subtend wall-following intersect on a wall is at the wall 'out there', and if there is any 'representation' of walls, it is only that produced by the robot's actual behaviour. Its route is a kind of mapping of the office environment for which it was built, a representation of that environment produced by the activity patterns resulting from its low-level reflexes.

We perceive the robot as following walls; the robot itself has no such aim or purpose - though, of course, it *was* designed for that purpose. The purpose is not represented within the robot, not used by the robot in the form of some internally represented goal, but because it was designed for that purpose, there can be no objection to saying that the robot *is* following walls. Its emergent activity was intended by its designer. It isn't just an arbitrary interpretation imposed upon that activity by us, the observers.

Establishing the legitimacy of a particular description of naturally occurring patterns of activity resulting from evolution is clearly not so simple. We cannot appeal to the designer's aims; the evolved behaviour does not merely emerge from lower-level mechanisms, it also emerged in the historical sense from the processes of genetic mutation, sexual reproduction and environmental selection. Is there any hope of isolating legitimate natural kinds in the behaviour of biological creatures? Can there ever be any justification, grounded in natural selection, for preferring one description over any number of others? I believe that Mataric's implementation of the higher levels of activity in her robot proves there can be.

3 Landmark Detection, Map-making and Navigation

The robot's structured activity produced by low-level reflexes in interaction with the particular environment for which it was built plays a crucial role in Mataric's implementation of the higher levels of landmark detection, map making and navigation. Mataric accepts that finding one's way in the world, in the sense of being able to retrace one's path and recognise routes which have been travelled before and detect revealing relations between different routes, requires some form of 'cognitive map', which may be seen as an internal representation of the environment. However, she maintains that this does not need to be an analytical or centralised representation.

Her first move away from abstract representations is to propose that the environment be coded in terms of 'landmarks' rather than distances and directions defined in terms of some global coordinate system. A landmark is generally conceived as an internally held representation corresponding to an objective entity in the world which serves as a point of reference. Space is represented as a collection of such landmarks and their relative locations. As long as the creature is always in sight of some landmark, and has a record of the sequence of landmarks it passed to get to its present location, it should be able to 'tell where it is'.

The advantages of landmarks over a more global representation in terms of spatial coordinates are that distances do not need to be precise and directions are determined by the location of landmarks along specific routes. A creature can only have landmarks which are located along routes it has physically travelled in the past, and their relationship to each other must be defined in terms of those routes.

Mataric takes this one step further. Her landmarks are defined in terms of the robot's own behaviour. The robot must continue heading roughly in the same compass direction for a minimum period of time, and at the same time the sonars must continue to report approximately the same readings, before a landmark of a particular type is recorded. There are four basic landmark types: left walls (LW), right walls (RW), corridors (C) and long irregular boundaries (I). Clearly these descriptions refer to objective features in the world, and clearly there has to be some reliable correspondence, in order that the robot may negotiate its office environment successfully, between its landmarks and the distinctive features of that world, but it would be wrong to conclude that the landmarks are *defined in terms* of objective features in that world.

The landmarks emerge from the robot's wall-following behaviour. It is only the fact that the robot can be relied upon to find and follow walls which gives substance to the four types of landmark chosen by Mataric. The algorithm to determine the type and bearing of a landmark uses a running average to eliminate spurious sonar and compass readings. The office environment is tolerant of this approach, since it consists mainly of right angles and straight walls. Wall-following, emergent from low-level reflexes, does not result in straight lines, but it does have a recognisable structure at the level of running averages. A creature adapted to a different type of environment would need alternative activity patterns to produce temporally structured inputs that would allow it to cope with the limitations of its sensors.

The landmarks are stored in the nodes of a distributed graph. The structure of this graph reflects the constraints on the robot's low-level behaviour. Wall-following restricts it to what is essentially a one- dimensional trajectory, with an occasional crossing of its own past route. This can be represented as a straight-forward linear list, in which the crossings are handled by jumper links to a switch-board. Each node in the graph is connected to its neighbours. Initially the nodes are all empty, and exploring the environment consists in filling the empty nodes with landmarks discovered along the route.

The first landmark detected in accordance with the algorithms described above is automatically stored in the first node (which is special, since it 'knows' that it is the first node). The node is then 'activated' to indicate the robot's current position in the graph. Whenever a subsequent landmark is detected, its type and compass bearing are broadcast to the entire graph. Each node compares that information with its own type and compass bearing. If none of the nodes reports a match, the landmark is assumed to be new and stored in the successor node to the one currently activated. Since the matching is done in parallel, map localisation is effectively achieved in constant time, regardless of the size of the graph.

The system is made more robust by insisting that landmarks can only occur in the context of its route (which is constrained by its wall- following behaviour), as well as by making the landmarks themselves into dynamic, time-averaging entities. Very little information about the environment is actually stored in the graph. There is no accurate representation of the landmarks' locations in Cartesian space. 'Facts' like the existence of corners, doorways and other types of convexity and concavity, the nature of furniture, and the characteristics of human beings who cross the robot's path, are all taken care of by the autonomy of the low-level reflexes. These facts do not have the power to confuse the higher level navigation system, since this relies implicitly on the independence and robustness of wall-following. There is thus no need to store such facts, or to plan for contingencies which might arise as a result of their existence. It is not necessary for the navigating layer to 'know' about environmental contingencies that the low-level behaviour is adapted to cope with.

On the other hand, the way Mataric's system operates is not a matter of a higher level planner 'calling' lower-level modules as and when they are required. The lower level reflexes are in *continuous* operation and decide *for themselves* when they are required. The robot can 'survive' perfectly well in its environment by using low-level behaviour alone. This allows it to construct its map by exploring the environment without any plan or goal. Even if the resulting landmark graph might be called a plan for action in its environment, it must be seen as emerging from this autonomous, low-level behaviour, rather than being imposed on it or directing it.

Mataric's map-making and navigation system takes Brooks's ideas several steps further. The important points, for the purposes of this discussion, are the following:

- Landmarks come into being as the result of the creature's structured activity within a particular environment. This structured activity allows the robot to detect correlations between its own movement and its sensory input and permits the categorisation of landmarks into types.

- Landmarks have extension over time. The algorithm for detecting them uses a running average which has to be maintained within certain values over a minimum period. There is no 'space' between landmarks in the landmark graph. In a sense, the robot's map chops the routes it travels into segments of activity rather than signposting them with positional landmarks. Navigation becomes a matter, not so much of travelling from one landmark to another, but of slipping from the behaviour specified by one landmark into the behaviour specified by the next.

- The notions of path and neighbour thus become very important. They allow spatial information to be encoded in terms of a succession of activity segments. Segments of activity stored contiguously (or at least in some way that allows 'expectation' to be passed from one to the other) perform a mapping of the environment in terms of travelled paths. The map and the paths do not exist inside the creature in any objective or global sense; they only come into existence when the creature is placed in the environment and allowed to move as dictated by its low-level reflexes and by the instructions in its landmark nodes.

- Neighbours and well-travelled paths provide 'context'. This context is initially established through wandering about aimlessly (but not randomly, since the wandering has a low-level structure imposed by wall- following) and covering as much as possible of the environment with paths. These paths map the environment in terms of segments of situated activity. Once this has been accomplished, information about the environment is always seen in context, because at any particular instant the robot will be doing something, and what it is doing is related to its location in the graph and to what it was doing at the previous location and will be doing at the next. No data ever occurs out of context, and the context does not require global knowledge about the environment or top-down specification in terms of type-hierarchies; it flows naturally from the robot's structured activity in the particular environment, which always ensures that the robot is firmly situated in that environment.

4 Mataric and Millikan: History of Use

Millikan's appeal to history of use is an attempt to establish a notion of proper function on evolutionary grounds alone. It depends on natural selection to confer its retrospective stamp of approval on her choice of natural kind. I argued at the beginning of this paper that the task of disentangling the contribution made by a (hypothetical) internal representation to the survival and reproductive success of a creature whose behaviour is assumed to depend on its use must be considered an impossible one. Evolution does not provide grounds for explanations based on formal task description or analogies with machines. This has led some authors (e.g. Gould and Lewontin, 1978) to conclude that a plausible adaptive story may be concocted to fit any explanatory hypothesis at all. How can natural selection alone justify the choice of one explanatory hypothesis, and its associated natural kinds, in preference to any number of others? What natural kind could ever be shown to have a functional history legitimised entirely by evolution, with no need for a quasi-teleological story in terms of a hierarchy of parts? I believe that Mataric's robot provides the outlines of a principled solution to this problem.

In Mataric's robot, the landmark detector and navigation layers rely on the structured, lower-level activity of wall-following. Wall-following is not a low-level 'module' called by these higher level activities as and when it is needed; it is in continuous operation and provides the behavioural and perceptual substrate which makes landmark detection and navigation possible. If this were a naturally evolved organism, the fact that it displays landmark-detection and navigational behaviour which depends so crucially on the existence of that lower-level structured activity would constitute evidence that wall-following must have been a stable element in the creature's behavioural repertoire for a considerable part of its evolutionary history. The use made of this dependable activity pattern in the evolution of a higher level of behaviour implies a stable functional history and identifies it as a natural kind.

Whatever changes might have taken place in the environment, no matter how complicated the genetic mutations and recombinations that went into producing the low-level pattern, and irrespective of the neuronal structure subtending the two behaviours, if it is possible to establish navigation by landmarks on empirical grounds, and if it is possible to establish on the same grounds that the particular landmarks used by the creature can only come about if the creature performs a certain type of structured low-level activity *which it does in fact reliably perform*, then that low-level structured activity acquires the distinction of a natural kind, and it acquires that distinction purely on evolutionary grounds.

The obvious advantage of explanatory hypotheses that are justified in this way is that all the elements are open to empirical scrutiny. Low-level activity patterns can be observed and related to the environment. Comparative studies with related species as well as controlled changes in the environment can be used to establish the criteria and parameters at both levels. The dependence of one upon the other can be investigated by artificially induced alterations in the low-level activity. This is very different from trying to justify, on ethological and evolutionary grounds, a model whose explanatory entities are assumed to be internal to the creature. The 'meaning' of an explanatory entity identified in the way I have proposed need not be sought 'in the creature's head', because the entity itself is not presumed to reside in its head. Nor does it depend on the observer's interpretation. Its 'meaning' is defined by the use that is made of it by higher levels of activity.

At the end of one of her papers (Mataric, 1991b), Mataric speculates on possible extensions to her robot. As a result of navigation by landmarks, the robot must produce a higher level of emergent activity. It may be possible to introduce yet another layer which monitors this activity in terms of some dynamic feature that results from its interaction with the environment. What Mataric appears to have in mind is a system that would evaluate and correct its own rules. I believe this to be wildly premature. The gap between such a system and the type of activity that Mataric's current robot is capable of seems too large to be bridged in a single step by a natural or artificial evolutionary process. In fact, I strongly suspect that even the step between wall-following and landmark navigation is unrealistically large, and that this accounts for the need to introduce physical parameters. (For a more gradual approach to a similar problem, which does not use such parameters, see Pfeifer & Verschure 1992.)

But I see no reason in principle why layers cannot continue to be grafted onto a functioning creature in the way that Mataric suggests. The important characteristics of the process are the emergence of a higher-level dynamic feature from the interaction between the robot's current level of activity and its environment, and the evolution of a sensing mechanism to track this feature. Each new layer will tend to be more abstract than the last, in the sense that it deals in 'features' which are further and further removed from information which could be obtained by examining either the environment or the creature's neurophysiology. The notions of 'navigation' and 'landmark' can be taken in a much wider sense than that connected with travelling through a landscape. One may think of an infant as learning to 'navigate' the space within its reach by the use of its hands and eyes, of a piano player as learning to 'navigate' the keyboard by performing situated patterns of activity in the form of scales (see Sudnow, 1978 for a somewhat irritating, but nonetheless illuminating personal account), and of children as using the environment provided by speakers of their native language and their inborn dispositions to produce certain structured vocal patterns to learn to 'navigate' the space of that language.

5 Conclusion

My aim in this paper has been to use Brooks's and Mataric's work on autonomous robots as a kind of existence proof for an alternative mode of explanation. No claim is made for the biological validity of their choice of activity patterns, since I believe that their approach is still basically one of top-down design, and therefore does not relieve us of the need to discover, through field-work and comparative studies, the particular natural kinds of human and animal behaviour. But I believe they have shown that:

- Patterns of activity whose high-level structure cannot be reduced to a specific sequence of movements may emerge from the interactions between simple reflexes and the particular environment to which they are adapted.

- Such patterns of activity can result in dynamic features with temporal extent which may correspond only very indirectly to 'features' in the world.

They owe their characteristics to the interaction between the structure of the low-level activity and the nature of the creature's environment. But their reliable occurrance could encourage sensors to evolve that would be capable of monitoring them, thus providing 'information' for the next level of activity.

- This type of configuration, if found in a naturally occurring creature, would confer the status of natural kind on the lower level of activity, since it would clearly establish a history of use for that activity in the creature's evolutionary past.

Acknowledgements

I would like to thank Margaret Boden, Dave Cliff, Inman Harvey and Maja Mataric for their helpful comments.

Reference

Brooks, R. A. (1992). Artificial life and real robots. In *Proceedings of the First European Conference on Artificial Life.*

Brooks, R. (1985). Robust layered control system for a mobile robot. Tech. rep. A.I. Memo 864, MIT A.I.Lab.

Brooks, R. (1991a). Challenges for complete creature architectures. In Meyer, J.-A., & Wilson, S. (Eds.), *From Animals to Animats: Proceedings of The First International Conference on Simulation of Adaptive Behavior*, pp. 434–433. MIT Press/Bradford Books, Cambridge, MA.

Brooks, R. (1991b). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Churchland, P. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* M.I.T. Press, Cambridge, Mass.

Dennett, D. (1975). *Content and Consciousness.* Routledge, Kegan Paul, London.

Dennett, D. (1978). *The Intentional Stance.* M.I.T. Press, Cambridge, Mass.

Dretske, F. (1988). *Explaining Behaviour: Reasons in a World of Causes.* M.I.T. Press, Cambridge, Mass.

Gould, J., & Lewontin, R. (1978). The spandrels of san marco and the panglossian paradigm: A critique of the adaptionist programme. *Proc. Royal Soc. London*, *205*, 581–598.

Harnad, S. (1989). The symbol grounding problem. *paper presented at CNLS Conf. on Emergent Computation, Los Alamos.*

Heil, J. (1981). Does cognitive psychology rest on a mistake?. *Mind*, *XC*.

Kuhn, T. (1970). *The Structure of Scientific Revolutions.* The University of Chicago Press, Chicago.

Mataric, M., & Brooks, R. (1990). Learning a distributed map representation based on navigation behaviors. In *Proceedings of 1990 USA-Japan Symposium on Flexible Automation, Kyoto, Japan.*

Mataric, M. (1991a). Behavioral synergy without explicit integration. In *Proceedings AAAI Spring Symposium on Integrated Intelligent Architectures, Sigart.*

Mataric, M. (1991b). Navigating with a rat brain: A neurobiologically inspired model for robot spatial representation. In Meyer, J.-A., & Wilson, S. (Eds.), *From Animals to Animats: Proceedings of The First International Conference on Simulation of Adaptive Behavior.* MIT Press/Bradford Books, Cambridge, MA.

Millikan, R. (1984). *Language, Thought and Other Biological Categories: New Foundations for Realism.* M.I.T. Press, Cambridge, Mass.

Millikan, R. (1989). Biosemantics. *Journal of Philosophy, 86.*

Nehmzow, U., & Smithers, T. (1991). Mapbuilding using self-organising networks in really useful robots. In Meyer, J.-A., & Wilson, S. (Eds.), *From Animals to Animats: Proceedings of The First International Conference on Simulation of Adaptive Behavior.* MIT Press/Bradford Books, Cambridge, MA.

Nehmzow, U., & Smithers, T. (1992). Using motor actions for location recognition. In Varela, F. J., & Bourgine, P. (Eds.), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.* MIT Press/Bradford Books, Cambridge, MA.

Pfeifer, R., & Verschure, P. (1992). Distributive adaptive control: A paradigm for autonomous agents. In Varela, F. J., & Bourgine, P. (Eds.), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.* MIT Press/Bradford Books, Cambridge, MA.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11.*

Sterelny, K. (1990). *The Representational Theory of Mind.* Basil Blackwell, Oxford.

Sudnow, D. (1978). *Ways of the Hand: The Organisation of Improvised Conduct.* Routledge, Kegan Paul, London.

Wilkes, K. (1989). Explanation — how not to miss the point. In Montefiore, A., & D.Noble (Eds.), *Goals, No Goals and Own Goals.* Unwin Hyman, London.