From S. F. Frowen (ed.), Hayek: Economist and Social Philosopher: A Critical Retrospect, London: Macmillan, 1997, 9-29.

2 The Connectionist Mind: A Study of Hayekian Psychology[1]

Barry Smith

INTRODUCTION

I shall begin with a discussion of recent work in cognitive science, and it may be useful to note that the title of this chapter might equally well have been something like 'Artificial Intelligence and the Free Market Order'. Readers might also care to note that I am, as far as the achievements and goals of research in artificial intelligence are concerned, something of a sceptic. My appeal to cognitive science in what follows is designed to serve clarificatory ends, and to raise new questions of a sort that will become clear as the chapter progresses.

Artificial intelligence (AI) research has two goals: that of simulating human intelligence via the construction of machines whose operations will in crucial respects be analogous to intelligent human performances,[2] and that of contributing thereby to our theoretical understanding of what human intelligence is. For the practitioners of AI the human mind is itself a sort of computer, and the structures and functions characteristic of the mind are in consequence seen as being capable of being realised in a range of different sorts of material, both organic and inorganic.

AI research has for some time been dominated by two competing methodologies, resting on two distinct paradigms as to what intelligence might be and what sort of computer might best be employed to simulate or reinstantiate it. On the one hand is the older and more orthodox 'symbolic' or 'symbol-processing' paradigm, which sees intelligence as a matter of the sequential manipulation of meaningful units (terms, concepts, ideas) of roughly the sort with which we are familiar in reasoned introspection. The upholders of this paradigm seek to construct artificially intelligent systems out of entities that are symbols (have both semantic and syntactic properties), operated upon in ways that model the rational processes studied by logic. On the other hand is the more modern, and in some respects more challenging, 'connectionist' or 'sub-symbolic' paradigm. This sees intelligence as a matter of the processing of units much more finely grained in character than those with which we normally suppose ourselves to be familiar in conscious experience. Such processing is to be conceived by analogy not with processes of reasoning as commonly understood, but with the massively parallel processing of electrical impulses by the billions of nerves distributed through the human brain, nerves bound together in networks that are subject to a constant and subtle forming and reforming of connections. In the corresponding simulations sub-symbolic entities participate in numerical – not symbolic – computation, and internerve connections are modelled by applying variable weightings to the numerical values (for example signal intensities) that pass through the system from node to node. The models here are derived from neuroscience, but also from statistical thermodynamics and other disciplines dealing with processes of an holistic sort.

PROPOSITIONAL VERSUS NON-PROPOSITIONAL KNOWLEDGE

For the defenders of the symbolic paradigm it is language, built up out of discrete and repeatable units, that provides the dominant theoretical model of cognition, and in his essay 'On the Proper Treatment of Connectionism' of 1987, still the definitive formulation of the connectionist paradigm, Paul Smolensky admits that the linguistic formalisation of knowledge brings with it important advantages. It makes the knowledge in question accessible to different people at different times and places, and it makes it possible for us reliably to check whether conclusions have been validly reached. It also helps to guarantee learnability: the beginner learns by following simple context-free rules that can be understood with little or no experience of the relevant domain. Science, accordingly, seeks in every case the sort of systematic clarity that linguistic formulation provides.

Connectionists, however, are interested neither in the rule-bound, deliberative efforts of the novice, nor in the sort of static knowledge that comes packaged in the form of scientific theories. Rather they are interested in the spontaneous, intuitive, tacit, dynamic, practical knowledge of the expert, and they would insist thereby that there are some spheres in which we might all claim to possess huge amounts of knowledge of this sort. Thus we are all experts, for example, in the business of everyday motion and perception, and this is also true of common sense and our native language.

Certainly the defenders of the symbolic paradigm can develop views of their own with regard to intuitive knowledge. Thus, for example, they can

embrace what we might call the Helmholtz hypothesis, to the effect that the intuitive processing of the expert is merely a case of fast and *unconscious* processing of the linguistically formalisable sort. Native speakers are seen from this perspective as unconsciously interpreting linguistic rules that could in principle be given explicit formulation in grammatical theories. Systems constructed on the basis of the Helmholtz hypothesis have, however, proved too brittle and too inflexible to model human expertise, and it has turned out that the task of establishing the needed articulations of expert knowledge in terms of context-free rules (for example for the domain of common sense) faces intractable difficulties of a hitherto unforeseen sort. As Smolensky would have it, the Helmholtz hypothesis 'has contributed essentially no insight into how knowledge is represented in the brain' (Smolensky, 1987, p. 7). Here, therefore, we shall concentrate primarily on the work of the connectionists.

A CONNECTIONIST MACHINE

It is useful to describe very briefly what a simple connectionist machine might look like (Figure 2.1).

The machine in question is a simple classificatory engine. Imagine a series of distinct numerical inputs (signals), derived, for example, from a battery of photosensitive devices trained in succession upon objects that it is the machine's job to recognise (classify). These signals, with given initial signal strengths, become transformed in their passage through the system according to the 'weights' of the links between successive nodes,

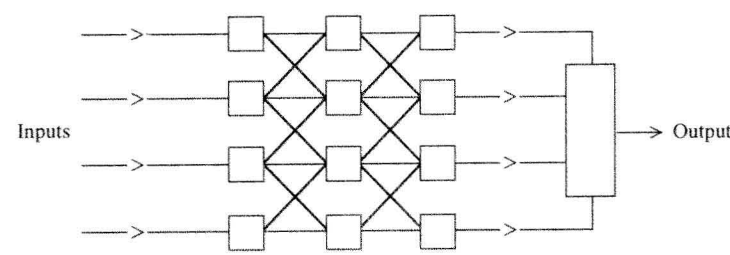

Figure 2.1 Simple connectionist machine weights that may be either positive or negative. The values passing through the machine are finally integrated into a single output, which is then associated with some value from a predetermined range of classificatory concepts (say: cat, rabbit, elephant and so on).

The behaviour of the classical ('symbolic') machine is fixed in advance by its program, so that the reaction of the machine to all admissible inputs is predictable from the start. The idea behind the connectionist machine, in contrast, is that it needs to be 'trained up' by means of a large number of initial trials, the results of which are controlled (in the simplest case) by a human operator who adjusts the relevant weightings by hand – and to some degree at random – in order to bring about improvements in the classificatory success that is achieved at each successive stage. If, on the basis of a given distribution of weights, the classificatory output is incorrect, algorithms can be used to adjust the weights of node linkages in such a way that the machine achieves a greater success rate in future trails. It is possible to prove that such adjustment will eventually, under certain reasonable sorts of initial conditions, reach one of a perhaps infinite range of stable states and that the machine in question will thereafter be able correctly to classify new objects from the given set without the need of further human intervention.

The idea of linguistic or propositional knowledge (of *knowing that*), which defenders of the symbolic paradigm take as their starting point, goes hand in hand with the idea of knowledge as a *memory store*, with which is associated a set of fixed, explicit rules for cognitive processing. The designers of connectionist systems, in contrast, set out to model the fact that *knowing how* is an evolving capacity of the knowing subject and is such that the content of knowledge and the process of gaining knowledge are not clearly separable from each other. There is no storehouse of memories from the connectionists' point of view; rather the connectionist system 'remembers' only in the sense that its processing patterns are subject to change, being constantly and cumulatively affected by what has been experienced in the way of processing in the past. The 'knowledge' in such a system is greater to a degree than under the symbolic paradigm, such as to influence the course of processing. Hence from the connectionist perspective it is necessary to reject the opposition between *competence* as a matter of knowing explicit rules and *performance* as a matter of the application of such rules.

In the world of standard symbolic AI (as here somewhat simplistically conceived), there is an initial plan (the program) controlling all the successive stages of processing in ways worked out in advance. Each of these successive stages is then intelligible on the symbolic level in virtue of the

fact that each unit (say 'cat' or 'horse') can be seen as enjoying an intelligible meaning of its own that remains fixed through time. In the parallel world of the connectionist machine, in contrast, knowledge and meaning come to be distributed among the nodes of the machine, so that, as Smolensky points out, what correspond to the concepts of which we are conscious are here 'complex patterns of activity distributed over many units. Each unit participates in many such patterns' (Smolensky, 1987, p. 8). The interactions between units are, it is true, entirely simple in nature: they are essentially a matter of numerical values being passed through the system and subjected to simple numerical adjustments at successive stages. It is our knowledge of certain numerical properties of these interactions that allows us, by adjusting weights in succession, stumblingly to bring the system to one or other stable state. But there is no way in which we can derive laws or principles that would make these processes intelligible at the level of concepts and meanings.[3] Considered from this perspective the individual nodes of the machine have no absolute values at all; their values are entirely a function of their position in the system of relations determined across the whole network.

Work in symbolic AI (again from our somewhat simplified perspective) has been marked by a tendency to produce simulations of intelligence that would apply only to rigidly well-organised worlds with a relatively small number of independent and well-behaved dimensions of variation. Such simulations have in addition proved 'fragile' in the sense that they have revealed themselves to be subject to almost instantaneous breakdown when an attempt is made to extend them to new and more complicated sorts of cases. Connectionist machines, in contrast, have been constructed that show a certain degree of elasticity. This, however, as we have seen, is at the price of intelligibility: the behaviour of the machine, while well understood mathematically, becomes from the point of view of our conceptual understanding a black box. From this, however, it follows too that the elasticity of even connectionist systems must be of a very limited sort. For if we do not understand, conceptually, what is going on inside the machine, then we have no means of reconfiguring it in systematic ways to take account of new circumstances or new sorts of problems. Suppose, for example, that we have a system we have taught to translate spoken into written English, a machine we have trained up in terms of standard English pronunciation. If, now, we want our machine to transcribe, say, English pronounced with a Welsh accent, then we have no way in which we can systematically adjust what has been learned so far in order to take account of this new problem, and we have to retrain the machine from scratch.[4]

Connectionism is in the eyes of some a radical position precisely because it is held to embody a view according to which intelligent, cognitive processes are themselves non-intelligible; that is, are such as to resist our attempts at theoretical understanding. This view has been most clearly expressed by Dreyfus and Dreyfus, two long-standing critics of symbolic artificial intelligence research, who affirm that the approach of classical AI 'seems to be failing because it is simply false to assume that there must be a theory of every domain. Neural network modelling, however, is not committed to this or any other philosophical assumption'. (Dreyfus and Dreyfus, 1988, p. 37) They conclude from this that any success of the connectionist approach will imply that one of the two primary goals of artificial intelligence research – namely that of contributing in positive ways to a theoretical understanding of intelligence as such – will have to be abandoned. We shall return later to the sort of knowledge we can have of mind and other 'complex phenomena'.

HAYEKIAN PSYCHOLOGY

The significance of the opposition between symbolic and connectionist approaches is not restricted to the field of artificial intelligence. It is employed by Dreyfus and Dreyfus also as the basis for a new sort of history of philosophy. According to the Dreyfusian scheme, philosophers are to be divided into two groups: *precursors of symbolic AI* on the one hand – above all Plato, Descartes, Brentano, Husserl, the early Wittgenstein – and *precursors of connectionism* on the other, among whom are to be included not least Heidegger, Merleau-Ponty, Michael Polanyi and the later Wittgenstein. From the point of view of philosophers in the first group (according to the Dreyfus account), human reasoning is a linear, propositional matter; it is constrained by context-free rules and built up out of sensory and conceptual 'elements' that are in the business of mirroring or picturing corresponding (predetermined) elements of external reality. From the point of view of the latter, in contrast, human reason is an intuitive, creative, contextual, holistic affair – a matter, again, of *knowing how* rather than of *knowing that*.

The philosophical story is of course a mite more complicated than Dreyfus would have us believe. Consider the philosopher-scientist Mach, a thinker who straddles the boundary between elementarism and holism in ways not readily allowed for in the Dreyfusian historiography. Mach's philosophy is centred around the idea that the world is built up of 'elements' subjected to constantly changing patterns of combination. The most important examples of such elements are the 'sensations' of standard empiricist philosophies. For Mach, however, sensations are classified differently from elements of other sorts, not because of what they are intrinsically, but entirely in reflection of their *relationship with other elements*: 'the elements of all complexes prove on closer inspection to be homogeneous' (Mach, 1959, p. 14). 'The antithesis between ego and world, between sensation (appearance) and thing,... vanishes, and we have simply to deal with the connexion of the elements..., of which this antithesis was only a partially appropriate and imperfect expression' (ibid.)

Mach's ideas will prove of interest in what follows also because he embraces a doctrine of what he calls the 'economy of thought': we are to understand the constantly shifting world of elements-in-connection as subject through and through to the law of least action; the world is a dynamic system within which elements, in forming connections, seek always the path of least resistance in such a way as to establish stable minima; and the behaviour of elements, including psychological elements, is for Mach to be understood entirely from this perspective (which Mach derives from evolutionary biology).

One thinker influenced tremendously by Mach is Hayek, and the doctrines at the heart of Hayek's psychology are in fact taken over directly from the older thinker. Hayek tells us that the problem of *The Sensory Order* is that of establishing the relationship between the 'two orders' of the subjective, sensory, perceptual and phenomenal on the one hand, and of the objective, scientific and physical on the other (Hayek, 1952, 110).[5] The main thesis of the work is indeed an ambitious extrapolation of Mach's own thesis concerning the nature and status of sensations. Hayek wants to show that *all* attributes of mental experience can be explained in terms of the system of connections of corresponding groups or patterns of nerve-excitations. He wants to show, in other words, that those mental properties with which we suppose ourselves to be acquainted through introspection, through the observation of other people's behaviour, through history, poetry and so on are – lock, stock and barrel – a matter of certain structural or relational properties of the nervous system. Where Mach, therefore, remains only on the threshold of connectionism in holding on to a view of mind as an order of *sensations* (and thus of proper 'mental' entities on the 'symbolic' level, to use Smolensky's terms), Hayek crosses this threshold by embracing a conception of mental properties as relational features of events occurring in the sub-symbolic domain of nerve-excitations. On the other hand, however, Hayek is less radical than Mach. For where Hayek places his faith in physics and the physicist's understanding of the nervous system, Mach sees the physical order and the associated notions of space, time and causality as themselves merely the spurious products of the 'economy of thought'.[6]

To see how Hayek's conception works we must note first of all that, while the external stimuli that cause activity in our nerve fibres can be classified in physical terms, stimuli that are physically the same will in many cases not appear to us as the same from a qualitative point of view. This is so, Hayek holds, because it is not *what* is transmitted along a nerve fibre that is responsible for its characteristic effects, but *that* the fibre is excited at all (127, 131, 142). More precisely, Hayek rejects views, such as those defended by Johannes Müller and Ewald Hering, according to which it must be some property of the individual impulses proceeding in the different fibres that in some sense 'corresponds' to the differences of the sensory qualities. Hayek maintains, rather, that the specific character of the effect of a particular impulse is due 'neither to the attributes of the stimulus which caused it, nor to the attributes of the impulse, but may be determined by the position in the structure of the nervous system of the fibre which carries the impulse' (135).

What we call 'mind' is for Hayek nothing else than 'a particular order of a set of events taking place in some organism and in some manner related to, but not identical with the physical order of events in the environment' (149). But how, then, can certain relationships between non-mental events turn them into mental events? This question can be approached first of all from an adaptive point of view: how is it possible that in a certain part of the physical order (namely in the organism) a sub-system is formed that has the property of intentionality in the sense that it can be said to reflect some features of the physical order as a whole, in ways that have adaptive significance? How does such reflection enable the organism to behave more appropriately in relation to its surroundings?

Hayek, in contract with his predecessor, is not concerned with questions of 'atomism' or 'elementarism'. Like Mach, however, he is a structuralist both with respect to sensory qualities and with respect to the world as a whole. Thus for example he affirms that 'the whole order of sensory qualities can be exhaustively described in terms of (or "consists of nothing but") all relationships existing between them' (155). Or again: 'the order of sensory qualities no less than the order of physical events is a relational order – even though to us, whose mind is the totality of the relations constituting that order, it may not appear as such' (156).[7]

There are no absolute qualities of sensations from Hayek's point of view. Indeed, belief in absolute qualities 'is probably one of the main roots of the belief in a peculiar mental substance' (191). Another such root is the belief in a storehouse of memories, which Hayek also rejects. Hayek also

anticipates the central thesis of AI research to the effect that mind (intelligence, cognition) can in principle be realised in a wide range of different material structures. As the very same pattern of movements can be instantiated in a swarm of flies and in a flock of birds, so the very same abstract pattern can be instantiated in a musical score and a gramophone record. And then, as Hayek points out:

> It is at least conceivable that the particular kind of order which we call mind might be built up from any one of several kinds of different elements – electrical, chemical or what not; all that is required is that by the simple relationship of being able to invoke each other in a certain order they correspond to the structure we call mind (229).

MENTAL TOPOLOGY

We are to imagine the mind as a receptor organ subjected to a constant barrage of physical stimuli. Some patterns of stimuli will tend to occur frequently together. This, Hayek hypothesises, will tend in turn to bring about especially close connections between the corresponding groups of fibres and it will tend to lead also to the formation of especially dense connections to corresponding central neurons. (Here Hayek defends a view very similar to that formulated by Donald Hebb in 1949, a view that even today serves to define one of the standard types of model used in connectionist research.)[8] It is in this way that impulses occurring in different fibres may come to be experienced as qualitatively equivalent. The nervous system thereby acquires a new sort of topological structure: a distance-function comes to be defined upon it, reflecting the degree to which different quality groups 'belong together' as a result of the bonds (or 'cell assemblies') between them.[9] The 'position' of a neuron in the resultant system is then a matter of connections, of the paths laid down to other neurons: 'groups of neurons may have a larger or smaller part of their connexions in common. We can thus speak of greater or smaller degrees of similarity of the position of the different neurons in the whole system of such connexions' (326). And then: 'A very high degree of similarity in the position of the different neurons in the system of connexions is likely to exist wherever the neurons are served by receptors sensitive to stimuli which always or almost always occur together' (328). Further, because we can expect receptors that are sensitive to the same kind of physical stimuli to be frequently excited at the same time,[10] it will follow that correspondingly dense networks of connections will also be established laterally between the corresponding central neurons.

In addition, the formation of cell assemblies will be similarly affected by changes in the inner states of the body (with the occurrence of feelings of pain, hunger and so on) as well as by corresponding *actions*. Thus we can take a step towards explaining the adaptive properties of the 'mind' as a system of relationships:

> in many instances it is likely that certain kinds of stimuli will usually act together on the organism when the organism itself is in a particular state of balance or of activity, either because the stimulus regularly occurs under conditions producing that state, or because it occurs periodically so as to coincide with some rhythm of the body. The impulses which register such external stimuli will then become connected with impulses received from the proprioceptors which register the different states of the organism itself (333).

A system of connections hereby grows up in which is recorded the relative frequency with which, in the history of the organism, the different groups of internal and external stimuli have acted together.

The psychological classification of physical stimuli is then effected, in the simplest case, as follows. Each separate primary impulse has, as a result of these acquired connections, a certain *following* (a train or wake) of secondary impulses (recall the passage of numerical values through the nodes of the connectionist system). It is the total or partial identity of this following that makes given primary impulses members of the same class (334).[11] Note that the followings are a matter of purely physiological connections between impulses – they are not a matter of *associations between (meaningful) mental events*. Here again, therefore, Hayek is at one with the connectionist tendency within contemporary cognitive science.

MULTIPLE CLASSIFICATION

The classificatory system hereby constituted is of great sophistication. Each impulse becomes a member 'not merely of one class but of as many distinct classes as will correspond, not only to the number of other impulses which constitute its following, but in addition also to the number of possible combinations (pairs, triples...) of such other impulses' (335). This generates a system of multiple classification of a sort in which

classificatory competence and classificatory performance are (as for the connectionists) one and the same. Moreover:

> the classificatory responses are not different in kind from, but are events of the same sort as, those which are the object of classification. In consequence it is possible that one and the same event may appear both as an object of classification and as an act of classification (336).

It is this internal nesting structure, Hayek holds, that explains how what is after all a mere system of impulses 'can produce "models" of extremely complex relationships between stimuli, and indeed can reproduce the order of any conceivable structure' (336). This in turn would explain not merely outer- and inner-directed intentionality (or in other words consciousness of world and of self), but also the unlimited range of mental comprehension (our capacity to direct ourselves intentionally to objects in all conceivable categories).

The sophistication of the system is increased still further via the distinction between effective and potential connections between neurons (two kinds of following). Hayek recognises also – with his connectionist successors – that there are influences transmitting *inhibition* (or in other words *negative weightings*), the existence of which 'extends the range of possible differences in the position which any one impulse may occupy in the whole system of connexions' by introducing the possibility of impulses having effects that are directly opposed to each other (346). The whole resultant process is then not so much a matter of 'classification' as of 'evaluation', since the processing engine is capable of making distinctions of degree (both positive and negative) as well as distinctions of kind (357).

The system of multiple classification can cope not only with isolated sensory inputs but also with complex Gestalt structures (higher-order categorial formations) of a range of different sorts:

> The fact that chains of further processes ('associations') can be evoked not only by the 'elementary' sensory qualities ... but also by certain 'abstract' attributes of different groups of sensations (such as figures, tunes, rhythms, or abstract concepts), has usually been regarded as an insurmountable obstacle to any physiological explanation of mental processes.[12] For the approach followed here no such difficulty arises: the problem of the equivalence of 'similar' complexes of stimuli is not different in principle from the problem why the same associations should become attached to different impulses which correspond to the same 'elementary' qualities (369).[13]

This sensitivity to Gestalt structures in the environment is built up only slowly, as the brain becomes tuned to those abstract patterns that are of adaptive significance. There is formed thereby a 'map' reproducing relations between classes of events in the environment in a process that is affected both by the sector of the world where the individual lives and by the individual's own body. This map (or 'mental model') is then of course not only very imperfect but also subject to continuous although very gradual change (526).

Its representations are structural (compare the imperfect and abstract way in which cartographical maps represent topographical features) and they are a matter of accumulated knowledge, so that the map represents only the kind of world the organism has experienced in the past. The map's job, from Hayek's perspective, is to facilitate a certain sort of cognitive laziness (or what Mach called 'thought economy'). It allows the organism to avoid the need to classify individual events and combination of events from scratch in each successive occasion of confrontation. The map's structural character may also help to explain a central feature of animal intelligence: plasticity – the capacity to reuse rules learned in one context in other structurally similar but perhaps otherwise quite alien contexts (a capacity that may also be illustrated, on the level of high intellect, in the use of analogy in science).

The map thus plays the role of *memory* in Hayek's theory – not however that of a passive memory *storehouse*, but of an active memory *competence*. Moreover this active competence in a certain sense precedes our articulate mental experience. For as Hayek points out:

> we do not first have sensations which are then preserved by memory, but it is as a result of physiological memory that the physiological impulses are converted into sensations. The connexions between the physiological elements are thus the primary phenomenon which creates the mental phenomena (250).

But now, because this memory competence cum classificatory order is subject to constant evolution, it follows that *so too is the sensory world in which we live*. The richness of this world, 'is not the starting point from which the mind derives abstraction, but the product of a great range of abstractions which the mind must possess in order to be capable of experiencing that richness of the particular' (Hayek, 1978, p. 43f). It is as if we create the world as we go along by imposing classificatory schemes upon the *materia prima* which is the physical substrate. The particular qualities we attribute to an object in our experience are in this sense 'not properties of that object at all, but a set of relations by which our nervous system classifies them' (637).

As was pointed out by John Gray (1986), there is a certain grain of Kantianism here,[14] though it is questionable whether this Kantian element (and the solipsism that threatens in its wake) can be seen as being extended through the whole of Hayek's thinking if the latter is to retain any degree of coherence. Note, too, that while the thesis that we create the world as we go along by imposing neurally determined classificatory demarcations upon physical stimuli may be capable of being defended in relation to the world of sensory qualities, it faces obvious difficulties when an attempt is made to extend it to the 'objective' worlds of physics and neurology. Fortunately there is nothing in Hayek's work to suggest that these domains, too, might be merely the reflections of evolved classificatory competences of physicists and neurologists.

Hayek's contention is that human beings in the course of their development build up a system of differentiations between stimuli in which each stimulus is given a definite place in a slowly changing 'objective' order of increasing complexity and sophistication. It is this place in the objective order that 'represents the significance which the occurrence of that stimulus in different combinations with other stimuli has for the organism' (217).

The central nervous system is an adaptive engine for the constant reclassification – on many levels (including conceptual and emotional levels) – of the legion of impulses proceeding in it at any moment. We create the world in which we live in the sense that there are, on the side of nerve excitations, no fixed conceptual units able to mirror or picture corresponding (predetermined) elements of external reality in a one-to-one way. Only insofar as the nervous system has learnt to treat a particular stimulus event as a member of a certain class of events, can this event be perceived at all, for only thus can it obtain a position in the system of sensory qualities.[15]

ARGUMENTS AGAINST HAYEKIAN–HEBBIAN CONNECTIONISM

First, one important argument against a view of the sort outlined above is of a type that has been marshalled against neurologically oriented work in cognitive science quite generally and relates to a characteristic misuse of language that consists in imputing to parts of the human organism what can properly – if words are to retain their standard meanings – be imputed only to the whole. This occurs, for example, in phrases such as 'the nervous system perceives', 'the system of acquired connections remembers', 'the processing engine learns', and so on. It seems not possible to specify the meanings of terms such as 'perceives', 'remembers' and 'learns' in these phrases in ways that will be consistent with our normal use and understanding of the terms (so that there is a sense in which we do not really know what these phrases mean).

Second, Hayek's view, like that of Ryle in *The Concept of Mind* (1949), has no way of dealing with deliberate, conscious thinking (with reasoning as a logical process) – this, too, is a standard failing of connectionist approaches – and its account of consciousness is correspondingly weak. Hayek himself treats 'conscious experience as merely a special instance of a more general phenomenon', holding that the sphere of mental events – that is, the sphere of events that are ordered on principles analogous to those revealed in introspection – 'evidently transcends the sphere of conscious events and there is no justification for the attitude frequently met that either identifies the two or even maintains that to speak of unconscious mental events is a contradiction in terms' (172f). This implies, however, that Hayek has, like Mach and his Gestaltist successors, no means of drawing a clear distinction between intentionality as a matter of reflection (or 'isomorphism') and intentionality as a matter of 'consciousness of' or 'aboutness' in the sense of Brentano and his followers.

Third, Hayek has no way of dealing with what we might call 'mental causality', or in other words with the connection (so central to the work of Ludwig von Mises) between reason, choice and action. The system of *The Sensory Order* leaves no apparent room for planning, for self-control and for the deliberate self-shaping of the conscious subject (no room, indeed, for any self or ego or for the unity of consciousness).[16] A conclusion of this sort might well be acceptable to the followers of Hume and Mach, for whom consciousness is in any case dispersed and reactive.[17] But what, then, is to serve as a basis for the methodological individualism that so pervades Hayek's work in the social sciences?

Fourth, Hayek's view has no way of explaining the relative stability over time of our cognitive faculties and our qualitative contents, and it has no way of explaining the massive similarities in cognitive capacities between one individual and another. Klüver inadvertently reveals the weakness of Hayek's position in this respect when, in his 'Introduction' to *The Sensory Order*, he praises Hayek's theory because it suggests certain definite lines of experimentation. 'For instance, it should be possible not only to change sensory qualities experimentally, but to create altogether new sensory qualities which have never been experienced before'.[18] The problem for Klüver and Hayek (as also for Hebb and his successors) is that it seems not to be possible to change or invent sensory qualities in this way: a certain *fixity of species* seems to pertain to the world of qualitative experience. The hypothesis of cognitive universals has indeed found empirical support across a wide range of sensory and intellectual phenomena in recent years, and it seems that most linguists and anthropologists would nowadays assume as a matter of course that the truth of this hypothesis is presupposed not merely by the intertranslatability of all known languages but already by the very fact of linguistic communication itself. Hayek's account of the acquisition of perceptual and conceptual classification skills seems moreover to amount in the end to a sort of *tabula rasa* view of these matters that seems difficult to square with what is now known about the impressive mental competences of new-born babies.

Fifth, it is not clear how an approach along Hayekian–Hebbian lines can do justice to the creative open-endedness that seems to be involved even in our day-to-day activities of mental classification (a feature of our mental life that Hayek is otherwise at pains to insist upon). For in the real world, as Wittgenstein, among others, has emphasised, classification is often far from being a neat and tidy affair, yet it is a limitation of standard connectionist systems that they can be made to operate (the corresponding algorithms of back-propagation can be applied) only if the relevant classificatory space is fixed in advance.[19]

THE THEORY OF COMPLEX PHENOMENA

What, then, can be learned in a positive sense from Hayek's work in psychology? Hayek shows, most importantly, that the central idea behind the connectionist paradigm is at home not only in psychology and neurology but also in the sphere of economics. For the mind, from the perspective of *The Sensory Order*, turns out to be a dynamic, relational affair that is in many respects analogous to a market process. The mind is a 'continuous stream of impulses, the significance of each and every contribution of which is determined by the place in the pattern of channels through which they flow', in such a way that the flow of representative neural impulses can be compared 'to a stock of capital being nourished by inputs and giving a continuous stream of outputs' (Hayek, 1982, p. 291).[20] In 'The Use of Knowledge in Society' (first published in 1945), Hayek describes the price system as a *mechanism for communicating information*; and then, in the mind as in the market system, it is remarkable how little explicit (conscious) knowledge is required by the agent in order for him to be able to react in appropriate ways to changes in his circumstances. In the mind as in the market the most essential information is passed on in the form of abbreviated 'signals' (as contextually situated *nerve impulses* or *prices*, respectively). The price system is

> a kind of machinery for registering change, or a system of telecommunications which enables individual products to watch merely the movement of a few pointers, as an engineer might watch the hands of a few dials, in order to adjust their activities to changes of which they may never know more than is reflected in the price movement. ... The marvel is that in a case like that of a scarcity of one raw material, without an order being issued, without more than perhaps a handful of people knowing the cause, tens of thousands of people whose identity could not be ascertained by months of investigation, are made to use the material or its products more sparingly; that is, they move in the right direction (Hayek, 1949, p. 87).

Both the mind and the system of the market order are products of thousands of years of evolution, and both evolved through a massive number of trials and errors and through associated processes of 'training up', as each successive generation of individuals learned on the one hand to be conscious, and on the other hand to play the role of market participant. However the analogy between mind and market breaks down at least in this respect: that the former cannot on pain of contradiction be seen as having come into being *ab initio* as the unintended consequence of intended actions.

The structural similarity between the psychological and the economic sphere is understandable further in terms of the fact that both have the same dynamic root, for economic change rests in no small part not on absolute values but on our (always relative) *valuations,* and the latter are a psychological matter, a matter of constantly shifting and changing networks of relations determined through and through by context and perspective and by the 'economy of thought'.

Minds and markets are also comparable in that the understanding we can have of each is of a quite different nature from that which we can have of physical systems. This is indeed, for our present purposes, the most important respect in which Hayek anticipates contemporary connectionism. As Hayek points out, the operations of the mind are non-perspicuous because we are not explicitly aware of the relationships between the different qualities that constitute the mental order: we merely manifest these relationships in the discriminations we perform (a case of *knowing how* rather than *knowing that*). Furthermore the number and

complexity of these relationships is so great, and they are subject to such continual variation, that in any case we could never reach the point where we could exhaustively describe them.

On the other hand the theory of connectionist devices – of Hebb Models and Boltzmann Machines, of Hopfield Nets and Multilayer Perceptrons – is mathematically well developed, and the question thus arises as to whether this new body of theory might not be used to throw new light on those 'complex phenomena' that are at the centre of Hayek's interests not only in psychology but also in economics and the social sciences in general. Does connectionist theory yield a new solution to the old problem at the heart of all economic theorising in the Austrian tradition, a problem (crudely put) that turns on the fact that the phenomena with which the Austrians wish to deal in their theory are seen by the Austrians themselves as being in a sense too 'complex' to be theorised about? Hayek himself, familiarly, holds that in relation to the complex phenomena of the social and psychological sciences we can have only 'qualitative' understanding, and not exact prediction. The complex phenomenon that is the economic system can thus not be made rational or subject to 'control'.[21] Indeed prediction and control are already impossible in biology (consider the problem of predicting how a given species will evolve), and social, psychological and biological phenomena are such that the notion of a law of nature as a relationship obtaining between a few phenomena, linked together by a simple relationship such as cause and effect – a notion most at home in the field of astronomy – cannot be applied. As Weimer puts it, 'the prejudice that in order to be scientific one must produce laws may yet prove to be one of the most harmful of methodological conceptions' (Weimer, 1982, p. 244). And as we are now only too aware, the prejudice that scientific knowledge in this sense has already been attained in relation to the social world has proved to be one of the most harmful concepts in the sphere of political economy.

Notes

1. This chapter is a result of work carried out on the project 'Formal-Ontological Foundations of Artificial Intelligence Research', sponsored by the Swiss National Foundation. I am grateful to Graham White, Wojciech Żetaniec, Gloria Zúñiga and participants in the Hayek Memorial Symposium for helpful comments.
2. Current chess-playing computers are not the result of research in artificial intelligence in this sense, since they operate on the basis of exhaustive search strategies quite different from those employed by human chess players.

3. The sub-symbolic programs of connectionism can, as is often pointed out, be translated in such a way that they can also be implemented on classical (von Neumann) machines; but the translated programs are then not the kind of 'symbolic' program that the Helmholtz hypothesis requires.
4. Partridge, 1990, p. 71.
5. References in this form are to the numbered sections in *The Sensory Order* (Hayek,1952), with the decimal point removed.
6. See Mulligan and Smith, 1988, pp. 150f.
7. As Hayek notes, a structuralist approach of this sort does, however, very well explain the existence of certain patterns of which we are aware. Thus, for example, it can explain the existence of those 'intermodal' attributes that are revealed in our pervasive use of adjectives such as strong, weak, mild, mellow, tingling sharp, thick, rough, bright, heavy, hot, gritty, corse, hollow, luscious, astringent, smooth and so on in relation to sensory qualities of every sort (158–163).
8. On the role of 'Hebb models' in the development of connectionism, see Cowan and Sharp (1988, pp. 88f). In a note to that paper it is also suggested that there is a link between Hayek's work and the 'Boltzmann models' of more recent connectionist research (ibid., p. 119, note 57). Hayek himself adverts to Hebb at various points in *The Sensory Order* (see p. vii, and also sections 249, 315, 334, 538).
9. The idea of mental topology can be conceived as a generalisation of the marginal approach adopted by Hayek and other members of the Austrian school in economic theory.
10. As Hayek writes: 'we shall expect fairly close connexions to be formed between the neurons served by neighbouring receptors which are sensitive to stimuli which occur frequently together because they emanate from the same physical objects, such as pressure and temperature, certain chemical agents acting simultaneously on mouth and nose, etc., etc.' (331).
11. In his 'The Sensory Order after 25 years', Hayek draws attention to the dispositional connotations of this terminology of following, and also to the redundancy and randomness involved in the workings of the neural classificatory system:

    The 'followings' of all the impulses proceeding in the central nervous system at any one time are thus assumed to determine the potential or readiness of the system to do new things – internally or externally. Which of these potential neural events (toward which the system is inclined at any particular moment) eventuates would be decided by the partial overlapping of these followings through which, by summation, the potential effects of those linkages would be made actual. Only where a sufficient number of impulses coverged on any one neuron would it be made to 'fire' and to send out impulses to hundreds or thousands of other neurons (Hayek, 1992, p. 290).

12. For example by Stout, 1915.
13. Mach, too, allowed his sensory elements to picture certain qualities of higher order. Cf. Schulzki, 1979, pp. 158ff.

14. See also Petitot and Smith, 1990, but contrast Smith, 1990, p. 265.
15. At one point Hayek goes so far as to assert that an event that sets up, via peripheral stimulation, impulses in the brain of an entirely new kind 'could not be perceived at all' (636).
16. As Brentano stresses (1973, Book II, Ch. IV), mere similarity and regular association, however complex and many-levelled, do not as such add up to unity.
17. For Mach, the self 'is not a monad isolated from the world, but a part of the world and caught up in its flow ... so that we will no longer be tempted to see the world as some *unknowable* something. We are then close enough to ourselves and sufficiently closely *related* to the other parts of the world as to be able to hope for genuine knowledge' (Mach, 1917, p. 462).
18. Klüver, 1952, pp. xxi f.
19. Compare the remark of Dreyfus and Dreyfus to the effect that 'All neural net modelers agree that for a net to be intelligent it must be able to generalize; that is, given sufficient examples of inputs associated with one particular output, it should associate further inputs of the same type with that same output. The question arises, however: What counts as the same type?' (Dreyfus and Dreyfus, 1988, pp. 38f). The opposition at issue here is neatly captured in the following remark by G. L. S. Shackle on what is, in effect, the classification problem faced each day in the business world:

    Kaleidic effects, typically the response of asset values to 'the news', the abrupt and necessarily unforeseeable reaction of expectation-formers to announcements which in their nature are unheralded and of a purport quite unknown in advance, pose important problems of notation. The basic and essential character of the kaleidic phenomenon renders inappropriate the accepted methods of analysis of *fully-known problems*. It is the fact, astonishing and yet natural, that in economics we habitually and, it seems, unthinkingly assume that the problem facing an economic subject, in especial a business man, is of the same kind as those set in examinations in mathematics, where the candidate unhesitatingly (and justly) takes it for granted that he has been given enough information to construe a satisfactory solution. Where, in real life, are we justified in assuming that we possess 'enough' information? (Shackle, 1972, p. 184).

20. Cf. also Lavoie *et al.*, 1990.
21. 'The rationality of complex systems is not localisable in a single locus of control, and it is therefore never "conscious"' (Weimer 1982, p. 245).

References

Brentano, Franz (1973) *Psychology from an Empirical Standpoint*, translated from the 1st German edition of 1874 (London: Routledge & Kegan Paul).

Cowan, Jack D. and David H. Sharp (1988) 'Neural Nets and Artificial Intelligence', *Daedalus*, vol. 117, no. 1, pp. 85–122.

Dreyfus, H. L. and S. E. Dreyfus (1988) 'Making a Mind versus Modeling the Brain. Artificial Intelligence Back at a Branchpoint', *Daedalus*, vol. 117, no. 1, pp. 15–43.

Gray, John (1986) *Hayek on Liberty*, 2nd edn (Oxford: Blackwell).

Hayek, F. A. (1945) 'The use of Knowledge in Society,' *American Economic Review*, vol. 35, pp. 519–530. Reprinted in Hayek, 1949.

Hayek, F. A. (1949) *Individualism and Economic Order* (London: Routledge & Kegan Paul).

Hayek, F. A. (1952) *The Sensory Order. An Inquiry into the Foundations of Theoretical Psychology* (London: Routledge & Kegan Paul).

Hayek, F. A. (1978) 'The Primacy of the Abstract', in *New Studies in Philosophy, Politics, Economics and the History of Ideas* (London and Henley: Routledge & Kegan Paul), pp. 35–49.

Hayek, F. A. (1982) '*The Sensory Order* After 25 years', in Weimer and Palermo, 1982, pp. 287–93.

Hebb, D. O. (1949) *The Organization of Behavior. A Neuropsychological Theory* (New York and London: John Wiley).

Klüver, H. (1952) 'Introduction', in Hayek, 1952, pp. xv–xxii.

Lavoie, D., H. Baetjer and W. Tulloh (1990) 'High-Tech Hayekians: Some Possible Research Topics in the Economics of Computation', *Market Process*, vol. 8, pp. 120–41.

Mach, Ernst (1917) *Erkenntnis und Irrtum: Skizzen zur Psychologie der Forschung*, 3rd edn (Leipzig: Barth).

Mach, Ernst (1959) *The Analysis of Sensations*, translated from the 1st German edition of 1886 (New York: Dover).

Mulligan, Kevin and Barry Smith (1988) 'Mach and Ehrenfels: The Foundations of Gestalt Theory', in B. Smith, (ed.), *Foundations of Gestalt Theory* (Munich and Vienna: Philosophia), pp. 124–57.

Partridge, D. (1990) 'Connectionism is Better for Engineers than for Scientists', in Tiles, *et al.*, pp. 55–84.

Petitot, Jean and Barry Smith (1990) 'New Foundations for Qualitative Physics' (with commentary by Adam Morton), in Tiles, *et al.*, pp. 231–61.

Ryle, Gilbert (1949) *The Concept of Mind* (London: Hutchinson).

Schulzki, Ewald (1979) *Der Mensch als Elementenkomplex und als denkökonomische Einheit. Zur Anthropologie Ernst Machs*, dissertation (University of Münster).

Shackle, G. L. S. (1972) *Epistemics & Economics. A Critique of Economic Doctrines* (Cambridge: Cambridge University Press).

Smith, Barry (1990) 'Aristotle, Menger, Mises: A Categorial Ontology for Economics', *History of Political Economy*, annual supplement to vol. 22, pp. 263–88.

Smolensky, Paul (1987) 'On the Proper Treatment of Connectionism' (University of Colorado at Boulder, Department of Computer Science, variously reprinted).

Stout, G. F. (1915) *Manual of Psychology*, 3rd edn (London: University Tutorial Press).

Tiles, J. E., G. T. McKee and C. G. Dean (eds) (1990) *Evolving Knowledge in Natural Science and Artificial Intelligence* (London: Pitman).

Weimer, Walter B. (1982) 'Hayek's Approach to the Problems of Complex Phenomena: An Introduction to the Theoretical Psychology of *The Sensory Order*', in Weimer and Palermo, pp. 241–85.

Weimer, W. B. and D. S. Palermo (eds) (1982) *Cognition and the Symbolic Processes*, vol. II (Hillsdale, NJ: Lawrence Erlbaum).