Realistic neural nets need to learn iconic representations

W. A. Phillips, P. J. B. Hancock, & L. S. Smith
Centre for Cognitive and Computational Neuroscience, Departments of Psychology and Computing Science, Stirling University, Stirling FK9 4LA, Scotland
Electronic mail: apaps@vaxa.sitr.ac.uk

The central problem that Hebb's neuropsychological theory (1949) attempted to solve was that of how animals learn to perceive. Despite continued empirical and theoretical effort, it remains unsolved. As Hanson and Burr (H&B) make clear, neural network theory now provides new perspectives on the general relation between learning and representation. The aspects they emphasise however do not seem to us to be those best suited to the task of understanding perceptual learning. Symptoms of the imbalance in the particular connectionist approach they develop include the following:

(1) In section 2.2 H&B suggest that one constraint on biological learning is that the input consists of incomplete and 'stingy bits of data.' The contrast between this and the evidence for a rich sensory data base as emphasized by Gibson (1966) and many others is not discussed.

(2) Section 5 shows that what the hidden units of nettalk discover can be described in terms of the phonological output required. There is no discussion of how significant structure that is actually in the graphic input can be discovered, however. In the reading task the decomposition of a word into letters is a major part of that structure but because that is a given in nettalk, it cannot be discovered by learning.

(3) H&B suggest that categorisation is the central function being analysed. nettalk does not categorise input letter strings however; it generates a phonetic structure given a letter sequence. The assumption that categorisation is the central task facing cognitive systems, although common, may be misleading. The primary task facing perceptual learning may be better conceived of as that of discovering useful descriptors that can be applied to novel objects within a domain as well as to familiar ones.

(4) It is assumed that the goal of mapping layers is to construct sets of internal symbols that are independent of each other. As a consequence, nets using coarse coding and those with internal structure within layers such as topological maps are neglected. These symptoms can be seen as arising from the central belief that all representation in neural nets is symbolic. We agree that symbols are powerful and important, but we also agree with Searle (1980) that they cannot adequately mediate interaction with the external world. For that, iconic representations are required. Differences between these two kinds of representation that are relevant to their neural implementation include the following (Phillips 1989). Icons or images have internal structures that are derived from the internal structures of their referents, preserve similarity relations, and support meaningful generalisation. The internal structures of symbols are arbitrarily related to their referents and neither preserve similarity relations nor support generalisation. Images are created within descriptive media that are specific to particular domains. Symbols are general purpose. Finally, images include analogue, nontategorical descriptors; symbols are categorial because their internal structure provides no basis for generalisation within the referential domain.

We are well aware that these are contentious and difficult issues, but do not see how they can be avoided in a discussion of representation and learning in neural systems. To reduce misunderstanding we have found it useful to refer to René Magritte's painting of a pipe with the words, "ceci n'est pas une pipe," inscribed beneath. This draws attention to three kinds of things in the world: pipes, pictures of pipes, and the word pipe. Our hypothesis is that just as there are two kinds of representa-

The analysis of the learning needs to be deeper

John E. Rager
Department of Mathematics and Computer Science, Amherst College, Amherst, MA 01002
Electronic mail: jereger@amherst.bitnet

Hanson & Burr (H&B) have written an intriguing response to the anticonnectionist school (or more accurately the "connectionism is not anything different school"). They propose that connectionism "can be used to explore systematically the complex interaction between learning and representation." It is the fervent hope of all connectionist researchers that this is true, and certainly much of the current interest in connectionism is because of the possibility that nets can learn in interesting, nonobvious ways. The analytic techniques and analyses presented here do not conclusively demonstrate this, however, and I think they, or their immediate extensions, are unlikely to do this.

First, breaking up the input space geometrically (section 4.2 ff.) works best in networks without recurrence, since the dynamics are simplest in such networks. When recurrence occurs,
as it does in much important research, we can think of the input pattern as including the (intermediate) output pattern and changing with time, until the network settles (or we give up). Although it is still possible to examine the input-to-output mapping surface, it is more difficult. It is also less valuable, because just knowing the surface ignores the dynamic process, which may be the important part of what the network has learned!

Suppose that the response is correct and connectionism does provide a way to study learning and representation. It will be necessary to develop techniques to do far more than study feature-like representation, as done here. Methods of detecting complex, structured representations and processes will be needed. Consider some of the structures found in connectionist networks:

(1) Winner-take-all subnetworks (Feldman & Ballard 1982) and various other competitive structures. In addition to being one way to implement variables, these are the main way to allow competing hypotheses in situations where the ‘answer’ is underdetermined or where conflicting evidence exists. (For example, the interactive reading model of McClelland 1986.)

(2) Subnetworks in which sets of connections all have the same, or functionally related values, for some principled reason. (For example, positional invariance in phonology; Toretsky & Wheeler 1989)

(3) The “minicolumns” of Strong and Whitehead (1989). If it is possible for connectionist networks to learn, ab initio, interesting “representations,” then we will need complex tools and techniques to unearth them. We will return to this difficulty in a moment.

It may not be possible for a network to learn complex structures like these without some initial internal structure. This relates to the “poverty of the stimulus” problem discussed briefly in section 2.2. I am puzzled by H&B’s apparent dismissal of this. (“In reality, however, the linguistic environments and data a learner encounters will be large and diverse.”) Although the arguments for needing a prewired internal structure (e.g., Chomsky 1986), are not completely convincing, they cannot be dismissed out of hand, especially since if they are true, they have drastic consequences for what an amorphous net would be likely to learn (or not learn).

One of the difficulties that have to be faced in the analysis of nets which have learned is that of units with multiple functions. Gorman and Sejnowski comment:

Although it is attractive to think of a hidden unit as a feature extractor, this may not be the best way to characterize a hidden unit’s coding strategy. As we demonstrated, the hidden unit is capable of encoding multiple features and even multiple strategies simultaneously. This kind of pattern encoding . . . is more suggestive of a model-based approach rather than simple feature extraction. (Gorman & Sejnowski 1988, p. 88)

This overlapping functionality is difficult to decode in a system which is mostly structured as a simple feature extractor. Imagine how hard it would be to find and describe correctly the functioning of overlapping winner-take-all networks or two (or more) overlapping columnar structures, only one of which is “in use” at a time.

Is it this ability to overload units, and the ability to respond with different strategies for different contexts, that sets connectionist learning apart? Some twentieth-century philosophers (e.g., Wittgenstein) assert that symbols do not exist outside their context. Maybe it is the ability of networks to easily (?) redesign their mapping from input to internal marker that is important. If this is true, any attempt to “translate” a network into a symbolic description will miss the point.

In summary, the idea that connectionism provides a laboratory for learning is a good one, but unless ways are found to extract processes and complex structures, it will be hard to use it for anything but proofs that some task is learnable in some way.

There is more to learning than meets the eye (or ear)

Noel E. Sharkey

Centre for Connection Science, Department of Computer Science, University of Exeter, Exeter EX4 4PT, Devon, England

Electronic mail: noel@cs.exeter.ac.uk

Hanson & Burr (H&B) are to be applauded for their analysis of the relationship between the back-propagation learning algorithm and the representations that it creates. An important contribution of their target article is its comparison of different techniques for analysing a hidden layer in terms of its “receptive” and “projective” representations. Although most of this material has been around for some time, it has been scattered throughout the literature and has not been put together in this form before.

H&B’s goals are more ambitious than only showing the utility of analytic techniques, however. They wish to demonstrate a methodology for studying the integration of learning and representation, a task which, they say, has not been accomplished by the combined resources of cognitive psychology, behavourism, and AI. At first blush, these seem like reasonable goals. After all, connectionists spend their time “teaching” systems adequate representations for a variety of tasks. But then one begins to ask, What do they mean by learning? Are they discussing all learning or just a restricted class?

H&B state that, “it is important to keep in mind that we are interested in learning processes that could actually occur in brains” (sect. 2.1, para. 1). This is important because, unless we are dualists (or theory dualists), we would hold that all learning must occur in the nervous system (or at least the body). The neurobiological status of the back-propagation learning algorithms discussed by H&B is not altogether clear, however. Methods have been developed for automatically changing weight coefficients for input/output mappings (and intermediate layers) so that arbitrarily complex functions may be represented. These methods have been used to model a variety of neural circuits. Thus, it seems reasonable to claim that these methods can be used to develop systems whose behaviour resembles the behaviour of an idealised neural circuit. The possibility that real neural circuits learn in this way has not been firmly established, however. As Churchland and Sejnowski (1989) point out, “back-propagation is biologically implausible, inasmuch as error signals cannot literally be propagated back down the very same axon the signal came up.” (p. 41)

Even if we accept an assumption of nearly biologically plausible learning, we must still ask about the scientific status of the two models H&B analyse in detail: PARSNIP and NERTALK. This question is not addressed in their target article. The analyses demonstrate localisation within the NERTALK model for vowel detection. But surely, if anything, such localisation militates against the neural plausibility of the model. (Losing one neuron could mean losing the ability to recognise vowels). This leaves a problem of determining the criteria by which the models are to be assessed (besides efficiency). Perhaps empirically testable cognitive criteria are appropriate here.

The point is that there is a gap between a model of the learning of some cognitive behaviour and the instantiation in a brain of the learning algorithm along with the representations developed for the task. Yes, we may be able to model both neural circuits and cognitive behaviour using the same learning algorithms, but it does not necessarily follow that we have demonstrated how neural circuits learn cognitive behaviours. What has actually been demonstrated is that the putative “brain-style” computation makes a very powerful metaphor for the learning of a class of cognitive phenomena: those that can be learned through the extraction of statistical regularities. The hope and indeed the promise of connectionism is that this relationship is more than metaphorical; that by developing
theories in which both cognitive and neural explanations are compatible, researchers will eventually establish the physical basis of cognitive functioning. Until then, however, this basis should not be an a priori assumption.

Finally, let us return to the question of whether the integration of learning and representation discussed in the target article is the kind of integration that psychology and AI need. As with most academic issues, the answer must be "yes" and "no." On the positive side, there is an enormous variety of learning tasks that require the extraction of regularities from the environment (internal and external). This has to be useful to AI, where large amounts of knowledge have to be coded into "intelligent" systems. The target article provides the AI researcher with a guide for analysing the effects of a learning algorithm on the representation it has created. More important, H&B briefly discuss how we can study the way learning develops "connectionist rules" in the form of consequent (generalisation) regions in the hidden feature space. It is a pity that more analyses were not conducted on the formation of generalisation, because this is one of the most important features of connectionist learning and representation.

On the negative side, the real challenge for connectionism (in AI and psychology) is not the learning and representation of arbitrarily complex functions using slow, incremental learning algorithms. It is the explanation of types of learning that require metaknowledge. When Kohler's apes (1925) succeeded in getting their bananas, they were putting together knowledge from disparate sources in a single sweep of insight. It is no use merely saying that this type of learning is an emergent property of exemplar based nets, or that such behaviour would naturally fall out of regions of generalisations.

What the target article shows is how representations for a particular task form into generalisation regions so that novel inputs with similar feature profiles will produce similar outputs on the same task. What has not been shown is how generalisations that have been learned in one task (or several) can be applied in learning a novel task or in solving a novel problem. This is surely the hallmark of adaptive intelligence. The learning of such metageneralisations will be difficult for connectionism until more is understood about the requirements of initial network structure and about multi-net learning.

Problems of extension, representation, and computational irreducibility

Patrick Suppes
Department of Philosophy, Stanford University, Stanford, CA 94305
Electronic mail: suppes@csl.stanford.edu

I am sympathetic with much of what Hanson & Burr (H&B) have to say about connectionist models. My comments concern problems that are central in my mind but not addressed in any detail by H&B. The first is the problem of extension critical to all kinds of learning. The second concerns what is meant by a representation, and what formal claim should be made about representations in relation to learning. The third concerns questions of computational irreducibility, suggested by H&B's own analysis of the two extended examples of Nettalk and PARSNIP.

Problems of extension. A central criticism of cognitive theories of mental processes that do not attempt to address questions of learning concerns how they can accommodate the learning of new rules or new procedures. Perhaps the classic example is the language development of a child. Certainly, the classical cognitive theories that deal with this matter have not had great success with any extended set of data. Perhaps even more important, there is really no serious principled account of how the child makes the extension and consolidation from one stage of learning a language to another. In theory, connectionist models seem to suggest more realistic procedures for approaching this central problem, but in the only extended examples given by H&B, namely, phonemic representation of English words (Nettalk), and learning to recognize strings of syntactic categories based on sentences from a large corpus of natural language (PARSNIP), there is no attempt at all to explain how the net would be changed as extensive new material was introduced. H&B make the good point that in both cases the two nets generalized rather well to novel material. This does not concern continuing correction for extensive additional material, but the question of how a net that has been trained will work on new material - an important question, but not as central in many ways as the problem of continued adjustment to extensive new material. PARSNIP, for example, has 585 nodes in the net consisting of 45 hidden, 270 input, and 270 output nodes. It would be of great interest to know what would happen as PARSNIP was extended from a thousand sentences to an easily collected and tested 10,000 sentences. Moreover, it would be especially desirable to reach some principled understanding of how this extension is made - although the primary question is whether it can be made at all, not whether we can understand how it is made (see the discussion of computational irreducibility below).

A second kind of extension involves the use of new and different domain of external input, paradigmatically the domain of exemplar-based nets, or that such behaviour would naturally fall out of regions of generalisations.

Problems of representation. I find myself uneasy at the vagueness of the concept of representation used by H&B, even though they have many useful detailed things to say about their approach to representation. What is missing is some formulation of the structural or formal theory of representation behind their analysis of what can and cannot be done by nets: the kinds of
Commentary/Hanson & Burr: Connectionist learning and representation

structural isomorphisms that exist between given structures and their representations. Typical mathematical examples include: Any Boolean algebra is isomorphic to an algebra of sets, or any regular grammar can be represented by a finite automaton in the sense of generating the same language.

Much of the debate between connectionists and cognitivists, who are skeptical of each other's ideas about representation or about learning, is vitiated by the absence of any very precise and agreed-upon characterizations of representation. It may be that the way to think about the representation given by a net is like the relationship between grammars of various types and automata of various types; but above all, the semantics must be brought into the picture for a satisfactory notion of representation.

To avoid misunderstanding on this point I want to emphasize that from the fact that universal computation in the sense of equivalence to a universal Turing machine can be represented by a net, it does not follow that problems of representation are properly understood in any mathematical detail. My own conjecture is that a detailed mathematical refinement of the various alternative notions of representation that can be used by connectionists will lead to a clarification and understanding of a variety of critical issues, especially those critical in the debate with cognitivists.

Computational irreducibility. My third problem concerns a detailed understanding of what the two examples, NETtalk and PARSNIP, are really doing. H&B present a number of interesting details, especially of a statistical kind, about the behavior of these two relatively large nets, but in some sense the statistical information given does not get to the heart of the matter; it does not offer a real understanding of how the nets are working. For example, it is not at all clear from the presentation to what extent the developments are ergodic or very strongly nonergodic. Is there strong path-dependence, with the development of the particular net configuration being very accidental in character, as opposed to being asymptotically ergodic and very much independent of the starting point or the path by which the results were reached?

I would conjecture that many features of the nets are accidental and thus do not represent something we would want to generalize about in any detailed way. This is not meant to be a criticism, but an attempt to bring out how to think about these nets. It is characteristic of probabilistic phenomena that we are lost if we try to understand their details in all aspects. It is rather like trying to attach importance to the actual sequence of flips of a coin who have a strong belief that the coin is a fair one and the trials are independent. But what we need to understand, then, is not so much the results (as in the case of the actual sequence of flips of the coin) but the process or the procedure, from a probabilistic standpoint, by which the nets are produced. What this suggests is that we should try to have a stochastic theory of the procedures rather than of the actual nets produced.

It is not clear to me to what extent we can have a satisfactory theory of the kind suggested. A pessimistic possibility is that the actual large nets produced in the future will be computationally irreducible, in that no analysis in terms smaller than the nets themselves will give anything like a really detailed and accurate account of how they work. It is one of the pessimistic outcomes of modern science that a variety of important phenomena have such computational irreducibility. We associate such ideas with random systems or deterministic chaotic systems. (For example, the modern conclusion is that we are not really going to be successful in predicting the weather for any substantial number of days in advance.) It may be that the behavior of the very large nets that will be constructed in the future cannot be predicted in advance either, and we can do no better than run them to determine an outcome.

If we scale up NETtalk or PARSNIP by two, five, or 10 orders of magnitude, it certainly does seem possible that we will be totally lost in the details and can only hope to have a kind of schematic understanding of how the nets work. To take this conjecture one step further, this could be a reason for pessimism about ever writing an adequate grammar of any natural language. The cognitivists can feel safe and secure with a few explicit rules here and there, but for the real details, especially once we add the many prosodic and other features of spoken language, it is hopeless to attempt a serious account. The net the brain uses for language processing and the language generated may be too complicated to ever be understood adequately.

Connectionist models: Too little too soon?

William Timberlake
Psychology Department, Indiana University, Bloomington, IN 47405
Electronic mail: timberla@ibacs.bitnet

In its relatively short life, connectionism has occupied the attention of theoreticians and researchers in areas ranging from the modeling of the neuronal relations in the olfactory cortex to the acquisition of speech and recognition. Contrary to the giddy assessments and limitless horizons that so often accompany the introduction of powerful tools, the world is not in imminent danger of being saved. It is equally obvious, though, that an infant connectionism is in no danger of dying in its crib, and, indeed, is already asking for the keys to the car. It seems likely that connectionism will continue to cut a wide swath through learning and behavior in both application and theory. Hanson & Burr's (H&B's) target article brings some welcome order to the diversity of approaches and claims, for those of us who are only busy bystanders.

One reason connectionism is so appealing is that the behavior of a simple network lacks predictability, not just in one variable stochastic sense, but in terms of the entire organization of inputs and outputs. Connectionist models often have a surprising life of their own, in this way resembling a living organism we can question. Unfortunately, how well a model copies the behavior of a particular living organism seems due as much to art as science at this point. In addition, it is frequently not clear whether this "self-organizing" unpredictability of connectionist models is due to a few powerful assumptions about processing and structure, or to actual relations to the determinants of behavior. For example, the recent identification of nodes with individual neurons seems doomed to failure on functional physiological grounds alone, except in restricted cases.

At the moment, the field seems driven too much by overinterpretations of the relation of the model to the phenomena without considering that a particular type of output may result from the way the model is structured or the input is presented. What is needed is a more formal analysis of the identification and typical behavior of different classes of connectionist models with different types of input. The taxonomy proposed by H&B may mark an initial step in such an analysis. The cited work on analyzing the effect of the number of hidden nodes also seems along these lines.

H&B's point is well taken that connectionist modeling provides an increased opportunity to study the relationship between learning and representation. Most traditional learning models fall in one cell or another of a four-cell table. Models focus either on representational issues, or on a simplified "path" or rule of learning. Models are also typically concerned with either the course of acquisition or the organization and absolute value of responding at stability. To some extent, connectionist models provide a type of model that can deal with both complex representation and simple paths, and with acquisition and asymptotic performance. They also provide the opportunity to include relatively complex sorts of constraints on the learning of...
the system. So far, though, not enough modelers have taken advantage of this type of flexibility and potential power in the approach.

It may be that the most important contribution of connectionist modeling is as a technique for helping us think differently. There is nothing like a clear example to help break down our assumptions about the kinds of models that can produce particular sequences or classifications of outputs. For example, I was surprised to learn that hierarchical representations of output are compatible with single-hidden-layer models. I began to wonder what a connectionist model of hierarchically controlled behavior patterns like a bird's song or courtship might look like. The potential applications of connectionist modeling are exceptionally varied. At the same time, the greatest challenge to its long-term contribution may well be providing an analytic basis for selecting a particular type of model and its characteristics, and relating the results to other types and levels of analysis.

Connectionist models learn what?

Timothy van Gelder
Department of Philosophy, Indiana University, Bloomington, IN 47405
Electronic mail: tvgelder@ucs.indiana.edu

Hanson & Burr's (H&B's) target article makes a valuable contribution to our understanding of connectionism in many ways, not the least of which is its impressive demonstration of the variety and power of analytical techniques that can, and perhaps should, be brought to bear in the study of networks. Their focus on the interaction of learning and representation makes explicit what has often been only an intuitive undercurrent of connectionist thought, and makes a compelling case for its systematic exploration.

Unfortunately, the call for recognizing and exploring this interaction is presented in the form of an implausibly strong claim to the effect that this alone is what makes connectionism a significant alternative to more traditional computational approaches. The authors are surely right that learning and representation are intimately related in connectionist models and that the very form of connectionist networks makes them particularly amenable to exploration by means of various sophisticated and revealing analytical techniques. Both features seem to be characteristic virtues of connectionism, yet this is perfectly consistent with the possibility that other aspects of connectionist models also constitute an important departure from more traditional models.

Thus, consider H&B's controversial claim that distributed representations "are not special in and of themselves and provide no new representational abilities that other knowledge representation schemes from AI or computer science do not have" (H&B, section 3). In context, this is rather surprising, for one benefit of the target article is its demonstration of the importance of a powerful new conceptual apparatus for thinking about connectionist representations. The concepts of partition boundaries, response surfaces, decision conics (up to and including such exoticia as hyperhyberboloids) and so forth stand in stark contrast with such familiar notions as primitive symbol, syntactic structure, and designation, and it seems on the surface highly unlikely that the former simply describe an implemental medium for the latter.

On what basis, then, do H&B claim that distributed representations "do not really differ from symbolic representations" (H&B, section 4.1)? In their taxonomy of connectionist representations, distribution is distinguished by the fact that "a single input causes many hidden units to become active" (H&B, section 4.1), rather than just a single hidden unit, as in localist representation. Activity in multiple units will be required when, roughly, the hidden units must classify the input patterns into categories which cannot be separated by a single "hyperplane" in the input space. In some more or less trivial sense, each hidden unit is activated in response to—and hence encodes—some "feature" of the input. Thus, a distributed representation amounts to a pattern of feature-encodings over the hidden...
units. This view is the crux of what turns out to be H&B's only argument, for they go on to point out that the use of "features" is common in various standard symbolic approaches.

Whether one can say that distributed representations differ from their symbolic counterparts depends, to some extent, on how the terms are defined, and so it is crucial to use definitions that slice the phenomena at their conceptual and empirical joints. H&B seem to be working with impoverished conceptions of both categories. Thus, it does not follow from the fact that symbolic approaches have often made use of basic "features" of the represented domain that any style of representation invoking such features simply amounts to some variant of symbolic representation. There is a variety of further formal conditions that need to be satisfied; in particular, we need the notion of a fixed primitive vocabulary of discrete symbols, grammatical-formation rules, composition of symbols to form structured expressions, and so forth (see e.g., Fodor & Pylyshyn 1988; Newell & Simon 1976; van Gelder 1990), all of which are conspicuous only by their absence in connectionist networks of the kind explored by H&B.

Conversely, distribution is far richer than the notion of multiple units actively encoding the current input, as opposed to a single unit. As H&B point out, all that is learned in a network is recorded in the weights, and it is essential to note that one set of weights simultaneously encodes all the transformations from the input to the hidden-level representations. Moreover, in a fully connected network each hidden unit's activity varies (sometimes only in a trivial way) as a function of the entire input, and in that sense each input unit's activity is encoded over the whole hidden pattern. These are just two ways in which distributed connectionist representations involve superposition of encoding responsibility. It turns out to be more fruitful in general to understand distribution in terms of superposition of representations rather than either multiple hidden units, or feature-based encoding (see e.g., Murdock 1979; van Gelder 1990); yet the property of superposition can be shown to conflict directly with essential aspects of symbolic representation, such as the role of "specific localized constituents" (H&B, section 4.1).

It is highly unlikely that there could be only one significant point of divergence between connectionism and more traditional symbolic approaches, so insisting that one has located the single "real" difference is apt to be self-defeating. If distributed representation is indeed in some deep way nonsymbolic, this would surely only enhance and strengthen Hanson & Burr's general claims about the significance of the novel learning and representation relation in connectionism. There is no need to attempt to make their point more interesting than they already are by attempting to squelch other quite congenial perspectives.

Connectionist learning and the challenge of real environments

Mark Weaver* and Stephen Kaplan†

*Department of Electrical Engineering and Computer Science,
†Department of Psychology, University of Michigan, Ann Arbor, MI 48109
Electronic mail: Mark_Weaver@um.cc.umich.edu and Stephen_Kaplan@umich.edu

Hanson and Burr (H&B) focus their analysis of connectionist learning, quite appropriately, on the development of "useful" representations. Their analysis seems to us to have limited generality, however. While their description indeed applies to the typical back-propagation (BP) nets, we doubt seriously that it applies to the majority of learning that occurs in natural systems. Since such systems must function in an environment that is inherently complex and uncertain, this learning process is necessarily constrained. In particular, it must meet the following challenges: (1) It must be able to develop representations without knowing how they will eventually be used and (2) It must develop representations that can serve a variety of purposes (as opposed to the single purpose of performing a specific input-output mapping). Thus, it is our contention that the general case of representation learning is more difficult and differs in nature from what occurs in the BP network.

Consider what occurs in a typical network of this kind. The network is trained to perform some mapping of a set of input representations to a set of output representations. In the process, a new hidden-layer representation of each input is developed. This new hidden-layer representation is a function of the mapping that the network is being taught. Furthermore, the hidden-unit representation is optimized for that mapping; the network may boil away any information in the input which is unnecessary. What this describes is a system in which special-purpose representations are developed within the confines of a tightly supervised learning procedure. This is characteristic of neither connectionist models as a whole nor of the challenges which face natural systems.

For natural systems, "perceptual" learning should be considered the general case. That is, a natural system is indeed faced with the task of forming "hidden-unit" (internal) representations of the objects in its environment, but it must do so in the general absence of immediate and explicit feedback about how those representations will be used. Without an explicit goal, the organism must develop representations that are general rather than special-purpose.

As we have pointed out (Kaplan et al. 1990), there is a body of connectionists work which addresses this more general representational learning problem, beginning with Hebb (1949) and including more recent examples such as Breitenberg (1984), Edelman (1987), Grossberg (1987), Kaplan & Kaplan (1982), and Palm (1982).

This perspective also leads to a reformulation of H&B's description of "the complex interaction between learning and representation." We would claim that the learning/representation interaction in natural systems is indirect and that it occurs as a part of the evolutionary process. That is, although the "content" of the representations an organism develops is surely learned, we suspect that the "form" is largely determined by its neural architecture. Evolution, in turn, must then discover an architecture which produces representations whose form is "useful." H&B demonstrate that BP networks are capable of tailoring their representations to the extent that whether a network develops "distributed" or "local" representations is contingent upon the mapping task at hand. We doubt seriously that natural systems are able to be so agnostic.

After years of difficult and complex debate, a serious attempt to sort out the real differences between connectionist and symbolic models and to discover the essence of what makes connectionist models unique is a brave undertaking. There are certainly encouraging signs in H&B's effort. They are clearly aware of the larger space of connectionist models, and they have identified some of the difficulties that natural environments present to cognitive organisms. We were therefore disappointed to see that these considerations have little impact on their central arguments. They have identified "the complex interaction between learning and representation" as the one truly unique property of connectionism. But as they have described this process, it applies only to the limited class of networks that they have analyzed, and furthermore, it is uncharacteristic of learning in natural systems. We do not doubt that the statistical methods that H&B and others (notably Elman 1989) have used will prove to be valuable tools in understanding the operation of BP networks, but we remain unconvinced that
they justify broad conclusions about the connectionist enterprise as a whole.

ACKNOWLEDGMENTS
We would like to thank Eric Chown for his valuable insights and suggestions.

Author’s Response

Learning and representation: Tensions at the interface

Steven José Hanson
Learning and Knowledge Acquisition Group, Siemens Research Center, Princeton, NJ 08540
Electronic mail: jose@tactatus.siemens.com

In reading the accompanying commentaries, one gets the feeling that connectionism or neural network research (if one prefers Toulouse) invites the same kind of subjectivity as a Rorschach test. Perhaps because of its early state of evolution or because of the many fields with which it intersects, connectionism seems to have many different identities. The commentaries provide a snapshot of the diversity of thought and the multiplicity of directions researchers are currently pursuing. There are several common themes and reactions to the target article, nonetheless, that we will attempt to summarize.

In the target article we focused on one aspect of connectionist modeling: how learning and representation in nets interact, with an emphasis on static rather than dynamic learning and representational properties. Some of what we discussed has already appeared in the literature in various forms, but we tried to provide an organizing framework in which other connectionist claims and tensions might be considered. These tensions include: distributed versus local representations, discrete versus analog computation, symbolic versus nonsymbolic representation, learning theory versus experiment, “scaled up” versus toy experiments, and technology versus science.¹

These contrasts may seem a bit motley, but one encounters them in any serious attempt to study how learning processes and representational structure interact. Nor are they all peculiar to connectionism; they provide continuity with the rest of the machine learning field. The first three contrasts are special to connectionist models. The sort of representation nets use may restrict their applicability to what they might represent. Thus, the distributed/local contrast is a major category for commentary on the nature of representation in nets. I argue below that the real issue is the advantages of a learning system with access to distributed representations. The second contrast (symbolic/nonsymbolic) follows naturally from the first; having accepted the advantages of distributed representations for learning, one might wonder what kinds of structures distributed representations are. Can they be expected to exhibit the locality, compositionality, and simplicity one normally expects of “symbols”? If not, what sorts of nonsymbolic structures can nets represent, and to what end? the third contrast (discrete/analog) concerns computation: What do nets do with their representations? How can they compute with them, and what? Analog computation is naturally related to both net operation and learning. The relation between analog and discrete computation involves deep questions about the kinds of algorithms and properties associated with each, as well as practical questions about memory access, organization and control associated with specific applications.

The second set of three contrasts is not specific to connectionist learning and representation yet it has come into some focus in the context of connectionism. The discussion of learning theory and experimentation shows how the simple assumptions in connectionist models that lead to successful practical results can generate theories. The relation between theory and experiment is complex, and although we stressed empirical issues, these should be examined in parallel from many theoretical frameworks. It is important at this early stage that theory and experiment (or application) have a meaningful long term dialogue and not become polarized. One of the most contentious theoretical problems concerns scaling, of which there are two types, one more crucial for experimental work and hardly ever discussed in theoretical contexts. We should keep in mind that nothing really “scales up” in a complexity/theoretic sense when it comes to learning. We must understand not only how learning systems apply their resources to a problem but also how they control them. The third contrast (technology/science) is a natural one for any computational modeling approach. AI has had similar problems for some time, in which computational technology is confused with science. It took AI some time to realize that not every working model implemented on a computer is a candidate explanation for cognitive processing. Nor does the criterion of usefulness really limit the number or diversity of computational explanations enough to provide a common set of principles about intelligence, artificial or otherwise. Explanations of complex natural phenomena are hard to come by. Connectionist models are committed to one particular set of simple assumptions that are coherent and continuous with many other computational approaches, allowing us to keep the science and technology straight.

In the following sections we will discuss each of these tensions in some detail. The explosion of this field in recent years makes comprehensive coverage of all possible models or approaches impossible; the target article accordingly passed over many interesting trends. This led some commentators (Rager; Kruschke) to complain that various aspects of representation or learning had been left out; other commentators (Langley; Brown & Oaksford; Lamberts & d’Ydewalle) objected that connectionist models have not studied such important phenomena relevant to human learning as planning and problem solving.

To the first type of complaint we would have to respond, yes, we did leave out many interesting issues concerning representation and learning. The geometrical account presented here would have to be augmented with at least a third dimension involving time to do justice to the dynamic properties of learning and representation. We do not believe, however, that recurrent networks
(even very realistic ones) will make the analysis of representation intractable (Bridgenman; Suppes, Rager), as their complexity increases they will become more complex to analyze. Explanations will be analogous to descriptions of the weather to someone not near the window. This is not to say that principles cannot emerge from complex probabilistic models. A simple homogeneous programming language for a large number of models will yield generalizations about underlying principles that may be relevant to many kinds of problem domains and may help to explain learning systems. Simpler recurrent nets have been studied (cf. Elman 1988) and provide reasonable examples of how such networks can be analyzed in the context of sequential tasks. The interpretation problem is complicated and requires some sort of trace of the representation in time.

The second type of complaint is not valid. A large body of work on planning, problem solving, and dynamical processing exists (to mention a few: Allen 1990; Dyer et al. 1988; Elman 1988; Jordan 1986b; Miyata 1988; Pearlmutter 1988; Pineda 1987; Williams & Zipser 1988); this is an area in which we might expect some significant progress over traditional rule-based approaches involving interactions among plans, schemata, and their realization as action. These models are quite recent, have had less visibility, and are more complicated to implement and to train with data. Such technical problems in the programming language, however, should be kept independent of the scientific issues that arise when applying various connectionist technology.

**Learning and representation**

Our thesis was quite simple. Let us try to clarify some more of its substance: Connectionist models are based on a simple set of assumptions about learning and representation. Learning consists of "simple" local rules and representation consists of "simple" processing elements. As suggested in our target article, a very large set of variations is possible, even models with traditional rule-based properties, as well as hybrid ones with features from connectionist technology and AI/symbolist technology. Part of the appeal of connectionism arises from a simplifying strategy. Symbolists who are familiar with powerful, extensible, recursive programming languages implemented in friendly programming environments on machines that have ample high quality cycles may see connectionist assumptions as somewhat restrictive. Connectionists are committed to facing difficult issues arising from the interaction of perception, learning, and memory with just "balls and arrows" as modeling tools. We agree strongly with Haberlandt and Philips et al., who point out that this sort of interaction arising from constraints on the form of the input representation is not trivial and that there are no obvious principles that will allow the generic design of connectionist models at this point.

Such constraints or "anchors" on a modeling language are not easy to come by. If one is allowed to adopt any formalism for representation and any sort of learning function with any sort of argument structure, one recreates the machine learning field, which has adopted, statistical, logical, and heuristic representations with either very local incremental learning rules updating probabilities or numeric weights and very global learning rules generalizing or specializing rules (Langley). If one is only allowed to make normative or optimality assumptions about the model one is led to mathematical statistics (Golden). Connectionism lies somewhere in a no man's land between these two extremes. Neither fish nor fowl, yet sharing many assumptions from the statistical area and exhibiting some of the diversity of the AI/machine learning area, connectionism provides a true alternative hypothesis about the interaction of learning and representation. As Timberlake points out, connectionism may show old problems in a new light and provide insights into ill-understood mechanisms.

**Distributed vs. local**

A strong undercurrent of many of the commentaries (van Gelder; Barash; Bridgenman; Hendler; Brown & Oaksford) concerned the nature of representation in connectionist models. This was one of the motivations for distinguishing connectionism from symbolism. Smolensky (1988) had coined the term "subsymbol" to make explicit the difference between local and distributed representation. Subsymbols provide the "stuff" that nets are made of and must apparently figure in accounts of how nets process information. Yet it is hard to point to unique properties of distributed networks solely in terms of representation. There are examples of a diversity of representational types in the actual nervous system (Barash), some diffuse and holistic, others narrow and local.

Various perceptual functions arise from coarse coding. Gaining resolution or acuity through coarse coding, however, is not quite the same as distributing a concept or representing knowledge. Suppose you distribute "your grandmother" into bits and pieces in a neural network. Without some independently defined set of operations that can modify, combine, or somehow transform your grandmother into some distributed form, it is just not clear what one gains from distributing her in the first place. One must somehow combine all the bits and pieces or one must define "distributed operators" that can do with a distributed concept something that cannot be done with a local concept. One way to interpret the thesis of the target article is that learning is one distributed operator that does provide for unique transformations of the concepts (using hidden units, for example) that cannot be accomplished with local representations. Again, a productive and unique property of nets appears to be the interaction of connectionist learning and distributed representations.

**Discrete vs. analog computation**

Many commentaries are concerned with the relation between symbolic (classical, discrete) computation and connectionist (analog) computation. Some commentators discuss formal properties of connectionist computation (Chater), while others seem concerned about whether connectionist models can be seen as symbol systems (Brown & Oaksford), or, given the distributed, analog processes inherent in nets, whether it is fruitful to consider discussing representation at all (Munsat). Maki pro-
vides some interesting discussion of how cognitivists and behaviorists are finding some common ground in connectionism. McCulloch and Pitts do have a theorem concerning nets with "circles," that is, recurrence. They write that: . . . every net if furnished with a tape, scanners connected to afferents, and suitable efferents to perform the necessary motor-operations, can compute only such numbers as can a Turing machine; second that each of the latter numbers can be computed by such a net and that nets with circles can be computed by such a net; and that nets with circles can compute, without scanners and a tape, some of the numbers the machine can, but no others, and not all of them. (McCulloch & Pitts 1943).

Note they are careful to distinguish between the finite state properties of the network and its ability to control a tape or external memory. Obviously, if one assumes an infinite tape, an unbounded stack or what have you, one can implement any sort of classical system. This is not very interesting. It is more difficult to provide types of memory access in the connectionist system without also assuming that, as in classical architectures, there is necessarily a distinction between external memory and internal control. We accordingly agree with Chater that the real challenge comes from providing access to memory states and memory organization in the connectionist hardware (avoiding distracting resource limitations inherent in both systems) without making the classical memory/control distinctions. It is still more challenging to specify the relation between memory organization and learning. We stand corrected by Chater: Function approximation abilities of networks are limited indeed to a large general but finite set of functions; we do not see this as much of a limitation, however, on what sort of relevant computation can be accomplished by a simple feedforward net -- certainly not in the same historically devastating sense as the perceptron's inability to implement functions that were not linearly separable.

Symbolic vs. nonsymbolic

As to symbolic representation, it is instructive to compare the learning rule from behavioristic theories of circa 1930–1950 with current connectionist models. One critical difference involves the variables in the learning function itself. For example, the learning rule in Hullian theory referred only to the behavior and the environmental event (reinforcement). Stimulus sampling theory (Estes 1959) actually came closest to specifying something that appeared in the learning rule as an internal stimulus. The stimulus sampling parameter provided an indirect measure of the rate at which the environment was affecting the internal state of the organism. Connectionist learning rules, on the other hand, provide a specific measure of the stimulus itself, as represented by the function (e.g., hyperplane) computed in the feature space. Learning acts on this internal representation, which refers directly to properties of the external world. So although we agree with Brown & Oaksford, that connectionist models are presently a far cry from being symbol systems, they do provide a unique opportunity to explore how representations can be used to predict and exploit the environment; they may thus delimit the sorts of symbol systems peculiar to brain-style computation and perhaps to biological systems.

Learning theory vs. experimentation

Computational learning theory has been experiencing a renaissance in the last few years, with an explosion of research relating mathematical statistics to computational complexity and showing polynomial time learning with positive and negative examples and even with positive only examples (Narayanan 1988; Shvaytser 1990; Valiant 1984). This arose partly because AI too has been turning to learning and partly because of the successes of nets on applied problems (e.g., speech recognition). Much of the recent work in computational learning theory has been formulated in terms of neural networks (Baum & Haussler 1988; Judd 1988; Blum & Rivest 1988; etc.); many of the proofs are therefore related to architectures, learning rules, sample statistics, and other specific features of neural networks, as well as to very general aspects of learning and generalization. So contrary, to some of the commentators' beliefs (Suppes; Levelt), computational learning theory is thriving in the context of connectionism and is providing new results that may be more relevant to "natural" learning systems.

In fact, much of the problem with the work on "identification learning in the limit" (e.g., Gold 1967) was that it was too stringent about the conditions and assumptions under which learning might occur. When one weakens these assumptions (as in, for example, "PAC" learning, Valiant 1984) or when one considers variations of "representation functions" (i.e., functions used for representing target functions for learning; Valiant 1984) then polynomial time learning is not an unusual result. This points to a more serious difficulty in relating learning theory to natural learning systems -- systems that may be successful in some restricted domain, or that are modeled upon some well known biological adaptation: It is not always clear how to frame theoretical questions to make contact with systems that do successful learning in some domain (see Baum 1989). Nets represent a restricted class of possible automata in which many interesting questions can be posed. This makes them useful to the theorist and amenable to theoretical analysis (contra Levelt and Suppes). We agree with Jordan that theory is crucial in the connectionist enterprise, but it seems equally clear that experimental learning work must interact with theory. Theoretical work cannot take the place of experimentation.

Toy experiments and "scaling up"

The problem of scaling arises in many contexts, usually in connection with intractability in the net's learning or processing speed when either the number of examples or the input dimensionality increases. Scaling problems may also manifest themselves as qualitative changes in the nature of the problem itself, not necessarily relating to size. For example, mapping text to phonemes is not the same as reading. Many commentators (e.g., Suppes) brought up this problem as one of extensibility. Others
cited similar scaling issues in the context of "real human learning," questioning the psychological and biological plausibility of connectionist learning systems. (Levitt, Lamberts & d'Ydewalle, Pavel, Sharkey, Weaver & Kaplan). The reply concerns the nature of the constraints on network modeling and where they come from. There is clearly both a technological question here (see next section) and a question about the heuristic value of nets in guiding further research and hypothesis-testing. Some commentators saw connectionist modeling as premature because it lacks theoretical foundations (Pavel); others found it less useful than rule-based approaches (Lamberts & d'Ydewalle) because "a compact description is not very informative in its own right." Weaver & Kaplan suggested that there must be a sensible balance between constraints and generality of learned representations: Nets must be able to develop representations without knowing how they eventually will be used and must serve a variety of purposes. Such complaints must be weighed against the competitors. If connectionist models are compared to rule-based accounts of the same phenomena it is not clear that theoretical foundations can be provided because much of connectionist theory must emerge from interaction with empirical psychological research. Pavel is right that it is early. Even so, AI has not done much at all in attempting to embrace such constraints in any meaningful way.

We attempted to demonstrate that a compact representation that performs a task has a good chance of being relatively singular and meaningful. Constraints coming from network architecture, learning rules or data from the world all make it likely that such compact representations are related to the given problem in some theoretically meaningful way. We agree with Weaver & Kaplan's agenda, but we think they may have missed the point. The fact that sigmoidal feedforward nets are "universal approximators" does mean that representations can serve a variety of purposes. Most interesting functions can be represented (by approximation). Weaver & Kaplan's first point is in fact the thesis of the target article: Developing representations without knowing how they will eventually be used is exactly the sort of tendency that emerges from nets with hidden units. The net adopts representations to perform one task (as far as it knows) and if "enough" constraints are present the network can apply the representation to new tasks. Weaver & Kaplan are basically right, however, that scaling this ability across various complex domains simultaneously remains an important challenge.

**Technology vs. science**

We were careful to point out that the emerging technology of connectionist models will (for a while) remain independent of cognitive and behavioral theories. We do not yet know how these models and their varied properties will map onto language learning, arithmetic learning or event perception, for example. Many creative connectionists are computer scientists, engineers, and mathematicians developing variations of these models that do not yet have obvious applications in cognitive or behavioral contexts. Our taxonomy was meant to illustrate the breadth of connectionist research, not to suggest that it is time to map these models into specific theories of cognition. The future applicability to substantive theories will depend on the productivity of the modeling language.

Many commentators seemed to think we were focusing on back-propagation. NETTALK and PARSNIP were indeed feed-forward hidden layer networks using gradient descent in delta rule error, but it has never been clear to us what exactly makes something backpropagation. Many theorists refer only to gradient descent, or to multilayer nets or to feed-forward nets as backpropagation. One must keep in mind that feed-forward net with hidden layers using the delta rule (setting parameter estimation aside for the moment) is the simplest sort of model that can do nontrivial function approximation. Variations of the model are easy to generate and usually work well. It is also worth noting that our taxonomy at least illustrates that many models are kissing cousins of backpropagation. When one hears (e.g. from Sharkey) that "back-propagation is biologically implausible" what is usually meant is that the "C-code" implementation is biologically implausible, not the implementation one might find in brains. We should be careful what we rule out with such platitudes or through failure of imagination.

Much can be said about the psychological plausibility of connectionist models and about net learning versus human learning. Levitt points out that "human learning displays characteristic features that are not captured by nets." The usual human learning features are mentioned: "one-trial" learning and adding new knowledge to old knowledge. Such comments are often prompted by hearsay or purported demonstrations of inadequacy of nets in modeling human memory or learning. One must realize that when there is a dissociation between technology and theory, it is easy to make bad models. There is no guarantee that nets will provide a good model of arithmetic learning, dyslexia, or any other problem. If one writes a bad program in LISP, one that doesn't model problem solving very well, one doesn't then indict the programming language. We don't know of any particularly persuasive principle of net construction that says it must "undo old knowledge to learn new knowledge"; in fact, I know of several cases to the contrary; nor does the anecdotal "one-trial" learning property seem to be a matter for much concern.

Such superficial criticisms underscore how rule-based approaches have lulled researchers into accepting convenient accounts of complex phenomena without dealing with the hard questions at the interface of learning and representation. We don't deny that "one-trial" learning and knowledge integration are important, and we can imagine several sorts of connectionist models in which one may be able to study such features of human memory. There are indeed important questions about how knowledge interacts in time and why in some cases it interferes and in others it does not. One-trial learning is also a dynamic knowledge property, perhaps illustrating that the growth of knowledge in time leaves opportunities for rapid acquisition. Most connectionists don't expect easy answers to these complex questions. The modeling language they have adopted places severe constraints on the kinds of phenomena they can currently account for. It is just this sort of conservatism that can lead to principles.
mechanisms, and explanations that may be particularly
equivalent to the tensions at the interface of learning and
representation.

References

Letters "a" and "c" appearing before authors' initials refer to target article and
response respectively.


Morristown, NJ: [rSJH]

Press. [GT]

Andersen, R. A. (1986) Value, variable, and coarse coding by posterior
parietal neurons. Behavioral and Brain Sciences 9:90–91. [SB]

prefrontal associational-projection cell populations in area 7a of the
macaque monkey: a study using retrogradely transported fluorescent

activated by both visual stimuli and oculomotor behavior. Experimental
Brain Research 67:316–22. [SB]

Distinctive features, categorical perception, and probability learning:
Some applications of a neural model. Psychological Review 84:413–
451. [aSJH]

Press. [aSJH]

cat striate cortex in behavioral choice. Psychological Research 50:223–
28. [BB]

Interactions of visual and motor-planning activities in the lateral intra-
parietal area (LIP). Society for Neuroscience Abstracts 15:1203. [SB]

view. Symposium on the interface: Statistics and computing science.
April 21–23, Boston, VA. [RMG]

Computation 1(2):201–07. [rSJH]

In: Advances in neural information processing systems 1, ed. D. Touretzky.
Morgan-Kaufman. [arSJH]

distributions by neural networks. In: Neural information processing
systems, ed. D. Anderson. American Institute of Physics. [aSJH]

Blum, A. & Rivest, R. (1989) Training a 3-node neural network is NP-
complete. In: Advances in neural information processing systems 1, ed.
Touretzky. Morgan-Kaufman. [arSJH]

Hall. [JKE]

Braitenberg, V. (1964) Vehicles: Experiments in synthetic psychology. MIT
Press. [MW]


Sciences 5:372–73. [WSM]

Bridgeman, B. (1980) Temporal response characteristics of cells in monkey
striate cortex measured with metacontrast masking and brightness
discrimination. Brain Research 196:347–64. [BB]

(1992) Multiplexing in single cells of the alert monkey’s visual cortex during
brightness discrimination. Neuropsychologia 20:33–42. [BB]

generation: The emergence of “equivalence relations” in neural networks.
In: Cybernetics and systems ’90, ed. B. Trogel. World Scientific
Publishing Corp. [GDAB]

(1990b) The development of “symbolic behaviour” in natural and artificial
neural networks. In: Parallel processing in neural systems and computers,
Publishers. [GDAB]

Burr, D. J. (1988) Experiments on neural net recognition of spoken and
written text. IEEE Transactions on Acoustics, Speech and Signal
Processing 36:7. [arSJH]

radon transform. Proceedings of the International Joint Conference on
Neural Networks. [aSJH]

abstractions: The message still counts more than the medium. Behavioral
and Brain Sciences 11:80–97. [KH]

Chater, N. & Oaksford, M. (1990) Autonomy, implementation and cognitive

units. In: Advances in neural information processing, ed. D. Touretzky,
Systems I. Morgan-Kaufman. [aSJH]

35:29–58. [aSJH]


computation. In: Neural connections, mental computation, ed. L. Nadel,
L. Cooper, P. Culicover & M. Harnish. Bradford Books/MIT
Press. [NES]

Cooper, P. (1962) The hypersphere in pattern recognition. Information and
Control 5:324–46. [aSJH]

Cover, T. M. (1965) Geometrical and Statistical Properties of linear
inequalities with applications to pattern recognition. IEEE Transactions on
Electronic Computers, vol. EC-14 3. [aSJH]

48. [CT]

Cybenko, G. (in press) Approximation by superposition of a sigmoidal
function. Mathematical Control Systems Signals. [aSJH]

Dertlech, M. (1987) A connectionist architecture for representing and
reasoning about structured knowledge. Research Report No. CMU-
BOLTZ-29. Department of Computer Science, Carnegie-Mellon
University. [NC]

Dietterich, T. G. & Shavlik, J. W., eds. (in press) Readings in machine
learning. Morgan Kaufmann. [PL]

behavior: Some reactions to parallel distributed processing. Journal of the
Experimental Analysis of Behavior 51:399–418. [WSM]

J. Wiley. [aSJH]

Behavioral analysis in theory and practice: Contributions and
Associates. [GDAB]

Dunn-Bankin (1933) Scaling methods. Erlbaum. [aSJH]

powerful and biologically plausible extension to backpropagation
networks. Neural Computation 1:133–42. [aSJH]

selection. Basic Books. [MW]

Technical Report 8801. Center for Research in Language, University of
California, San Diego, CA. [arSJH]

(1989) Representation and structure in connectionist models. Technical
Report 8900. Center for Research in Language, University of California,
San Diego, CA. [MW]

Brain Sciences 5:777–78. [WSM]

57:94–107. [rSJH]

Base-rate effects in category learning: A comparison of parallel network
and memory storage-retrieval models. Journal of Experimental
Psychology, Learning, Memory & Cognition 15:556–76. [JKE]


Science, Special Issue 9:1. [aSJH]


Feldman, J. D. & McKonick, K. B. (1989) An empirical comparison of ID3 and
back-propagation. Proceedings of the Eleventh International Joint
Conference on Artificial Intelligence, Detroit, MI. Morgan
Kauffman. [PL]

BEHAVIORAL AND BRAIN SCIENCES (1990) 13:3

515