A Class of Parallel Cognitive Systems That is Neither Connectionist Nor Symbolic

Richard P. W. Loosemore

Department of Psychology
University of Warwick
Coventry, CV4 7AL
United Kingdom
psraf@warwick.ac.uk

Abstract

Connectionism seems to have defined itself as being more restricted in scope since back-propagation attained its present dominant status, with the emphasis shifting away from constraint-satisfaction and towards the study of networks of simple "neuron-like" units. A broader class of systems was implied by the earlier position, and the main goal of this paper is to propose an outline definition of this class. Two general aims inform the work: the immediate practical one of addressing problems that stand in the way of a connectionist model of higher-level cognition, and the less tangible issue of what to do about cognition if it turns out to be a mathematically intractable emergent product of a deeply non-linear complex system. It is concluded that the proper study of cognition may require a systematic exploration of the properties of systems like those defined here.

Introduction

Here is connectionism stripped down to its essentials:

- Many aspects of the symbol-processing paradigm seem inadequate for a close analysis of cognition.
- This might be attributable to differences between von Neumann computers and the computational characteristics of the brain.
- Therefore try to build models which are more consistent with brain-style computing.

One thing that helped motivate the shift towards connectionism was the discovery that some psychologically interesting properties (including distributed representation and generalisation) could be observed in simulations of networks made from simple, neuron-like processing units (e.g. Hinton & Anderson, 1981; Hopfield, 1982; McClelland & Rumelhart, 1981).

The original motivations for doing connectionism were such as to define a fairly broad class of systems that might be thought to have interesting properties, but in practice the initial "example" systems (with all their shortcomings) seem to have become identified with the field as a whole. There is little justification for excluding from connectionism systems which have, for example, some amount of complex internal structure, or which communicate via signals that are more than just single-valued activation levels.

This paper presents an outline of what the more general class of systems might look like, together with an indication of how they might be used to overcome some of the problems that beset current connectionist models.

The study of these systems would be difficult because of their intrinsic complexity. The definition is therefore presented with a view to the systematic, parameterised exploration of the properties of such systems. The eventual goal of the research is to investigate the behaviour of a variety of these systems in order to discover relationships between global behavior and local mechanisms.

Complex Systems

One aim of this paper is to suggest that many researchers have given a wide berth to systems more sophisticated than simple neural nets because of a reluctance to face a somewhat threatening possibility: that complex, non-linear systems are likely to have global properties that are not predictable from a knowledge of their local behavior.

The term "non-linear" might originally have referred to relationships which could not be plotted as straight lines, but its current usage is a good deal more subtle. If it is possible to take a system to pieces, model the behavior of its parts, then understand the system as a whole by (in some sense) combining the models of the parts, then the system is linear. Most of the physical world is linear in this sense, but it is distressingly easy to devise computational systems in which knowledge of the parts does not illuminate the behavior of the whole.

These systems are "complex" in a deep sense: it may well be that no scientific or mathematical account can be found for some aspects of their global behaviour. All that can be done is to build simulations of them, and perhaps observe empirically that some local behaviors give rise to particular kinds of global behavior. Properties like this are "emergent" in a practical sense, if not in a metaphysical one.

It may never be possible to gauge the extent of this pessimistic scenario: at any time we may discover some powerful new mathematical tools which allow complex systems to become "analytic" in the way that analytic systems are now. However, there is little sign of such a development at present.

Early Connectionism

Interpretations may differ, but at the outset it seemed that connectionism put the emphasis on the exploration of systems which were driven by the need to satisfy "simultaneous mutual constraints" (McClelland, Rumelhart & Hinton, 1986). The rest of this section outlines some of the points of common purpose that defined that early attitude.

Levels of Description

The brain is clearly a parallel computational system at the neural level. Cognitive modelling in the "symbolic" paradigm, however, takes its inspiration not from any kind of parallel computer but from the von Neumann architecture, in which there is one processor, one large store of symbols, and the former manipulates the latter. Part of the credo of the symbolic school is that it doesn't matter what kind of machine you are using at the lowest level, because any computer with more than a modest amount of sophistication can be set up to emulate any other.

For a connectionist, claims about computational equivalence are not relevant in this context because the issue is about *appropriate* kinds of description. There are many ways to choose a computational formalism, each of which can be used to build a (technically equivalent) description of a given system. But some descriptions may be simpler than others. As part of the search for a simpler, more appropriate way to describe cognition, a connectionist makes the following assumption:

> *Descriptive Levels Continuity*
>
> The properties of a cognitive system at the neural level (simple units, massive parallelism, etc.) are likely to be reflected in [simple, powerful, appropriate] descriptive levels that lie above the neural level.

Observing the extreme discrepancy between von Neumann-inspired models and the neural substrate, the connectionist concludes that either (i) there is an intermediate level which has some characteristics of both, or (ii) there is a modification to the symbolic level which gives it some "neural" characteristics (and makes it more adequate), or (iii) both of these. Either way, exploring the properties of connectionist systems is a way to clarify the situation.

Notice, however, that nothing in this position forces a retrenchment to a "strictly" neural kind of model: it would be equally valid to look for an intermediate level of description at which the computational primitives had mixed symbolic/neural characteristics.

Distributed Processing

If a cognitive system looks, on the surface, to be behaving in a structured, ordered, controlled manner, this does not necessarily mean that there are explicit mechanisms inside responsible for each aspect of this control. It may well be that an apparent "executive" module is actually a collective effect of a number of widely distributed units which are not specialised for only this one function. Exactly the same argument applies to the storage of information in distributed representations. Both confer some robustness on the system if the processors are unreliable or liable to damage. Also, if a function *can* be distributed, the system might work more efficiently because of the intrinsic parallelism.

Multiple Weak Constraints

If a system is designed so that every important event in it is constrained by as many independent factors as possible, spurious constraints will most likely be overwhelmed by the majority trend. Also, in the absence of any decisive constraints, a collection of weak ones which nevertheless all push in the same direction can be just as effective.

Learning

Symbolic systems tend not to be able to learn. This is sometimes a consequence of the methodological point that it is easier to pick apart a given behavior than to say how it could have originated. Other times it may be a matter of a doctrinal commitment to the existence of innately specified structures. For a connectionist, the emphasis on learning has a lot to do with constraining the models of adult competence: a learning mechanism which reaches a particular adult state is harder to devise than a direct model of that state.

Neural-Network Connectionism

Connectionist research has always placed a high premium on demonstrations of network capabilities. The backpropagation algorithm, for example, exemplified the fact that networks with hidden layers could learn to associate pairs of patterns. There are a number of ways in which this emphasis on existence proofs may have inhibited the proper development of the field.

Simple Systems

There is a reluctance to give neural nets too much complexity, either in the way of node computations, link signals or overall architecture. This seems partly driven by ideas about neural plausibility, coupled with a simple set of ideas about what neurons actually do, and partly a matter of keeping a good distance from any hint of symbolic approach. There is also, perhaps, a feeling that neural nets are hard enough to understand as it is; that no research methodology is readily apparent to enable any construction or study of more complex systems.

Backpropagation

The neurally implausible requirements of backpropagation (that activation signals travel in one direction while error signals travel in the other) seem to be tolerated because this type of system is regarded as an exemplar, and not necessarily as a direct model of any biological system. That said, this kind of net still makes frequent and detailed appearances at the core of "models" of various cognitive functions (Seidenberg & McClelland, 1989; Rueckl, Cave & Kosslyn (1991); Elman 1990).

Part of the attraction of backprop is that a mathematical proof exists of its statistical convergence behavior (Rumelhart, Hinton & Williams, 1986). For many interesting complex systems, it is unlikely that proofs will ever be available for any of their interesting properties, so the fact that backprop does have one seems to have stifled the exploration of other kinds of system.

A more puzzling development is the fact that the simple recurrent net (Elman, 1990), which is a variation on the backprop idea, has entirely cut loose from the mathematical proof (it is actually not valid for SRNs).

Supervised Learning

Again, the success of backprop as a learning mechanism has allowed the implausibility of supervised learning to go largely unchallenged. It is arguable that any network that needs a "teacher" during the training phase is leaving most of the intelligent work to the teacher.

Generalisation

The kind of generalisation achieved by most neural nets is of a very weak kind: patterns are similar only to the extent that they overlap. If a network learns about some regularities, those regularities are never used as the basis for any further learning.

A Class of Parallel Cognitive Systems

Having reviewed some characteristics of both original and post-backprop connectionism, this section takes a more prescriptive line. The goal here is to try to give shape to a type of system that lies somewhere between the strictly neural level of description and the symbolic level. Or, perhaps this *is* the symbolic level, but modified to make it more neural. Many of the suggested mechanisms could only be properly tested by exhaustive simulations under a variety of conditions, so at this stage the aim is just to moot the possibility that these systems need to be investigated.

There is also here an attempt to be provocatively complex: suppose the real story of cognition were to be as complex and (especially) as full of interdependencies as implied here? Present approaches which favour specialised research paradigms for each individual micro-domain of the cognitive system would stand little chance of discovering any significant aspect of the design.

Granularity

Cognitive systems treat the world as if it were a changing *configuration* of more or less stable *things*. A "thing" is represented internally by a computational structure that is referred to here as an *element* [this neutral term is introduced so that "symbol" and "unit" - which are equally near to the intended sense - can be used for references to the equivalent constructs in the symbolic and connectionist paradigms, respectively].

Distributed Processing

Begin with the assumption that the system is homogenous: all elements are of the same type, although each adapts to represent something unique, and all regular behavior that can be observed in the system is the result of the interactions of elements with one another. In other words, try to localise all functions within elements, rather than assuming at any stage that there are specialised mechanisms elsewhere which operate on elements. This assumption is a point of departure, not a dogma: it should be qualified by a (cautious) sensitivity to empirical data about the brain.

Distributed Representations

A distributed representation, in the connectionist sense, divides the encoding of a "thing" across a number (possibly all) of the computational units of the system. To insist that units deal in "microfeatures," in this way, is to force the discussion down one level of description, whereas it may be possible to build models at a level at which elements (as defined above) are a coherent construct. These elements may at some later stage be realised as distributed structures, but since it makes an investigation of their properties much harder, assume for the moment that we are not forced to work at that level.

Active Elements

At any given moment there is a set of elements which are in the *active* state. These can be roughly divided into three groups: those representing the current state of the world, those encoding patterns of motor output that the system is currently engaged in, and those which express whatever abstract situations or plans the system is "thinking" about (which latter, notice, need not be directly determined by what it is experiencing or doing at the time).

Element Behavior

Active elements have connections to one another by which they exchange information. The *behavior* function of an element determines what it does with the information that comes along these links. In general terms it tries to constrain other elements so that the active set collectively represents the world in a way that is consistent with what the system has learned from experience. This is a very simplistic characterisation - from here on out the problem is to specify in more detail what kind of (i) connection mechanisms and (ii) element behavior could give rise to various aspects of high-level cognition.

New Connections and Elements

One problem with standard connectionism is that for high-level cognition you need to make rapid links between arbitrarily different concepts. If all nodes are connected to all others this is not a problem, but we know that this is an unrealistic thing to expect of real neurons. In connectionist nets total connectivity (at least between layers) is often assumed, but the weights in these total-connect architectures do not change fast enough to make rapid, transient linking a serious possibility (learning mechanisms almost invariably require slow changes).

There is a cluster of related difficulties that seem to follow from this rapid linking problem. Connectionist nets cannot bind or use variables, nor make an easy separation between general categories and particular instances. Multiple instantiations are especially thorny: how does a network which encodes "horse" as either a single node or a pattern of activation across many nodes react to the sight of two horses? These are core issues in any reconciliation of symbol-processing and connectionism.

If real neural networks were able to do more sophisticated things than we presently suppose, then it might be possible to deal with the hot linking problem. We think we have a rough idea what neurons are capable of, but this knowledge is by no means very deep or reliable (Crick). One of the stranger habits of standard connectionism is that having broadly delimited the primitives that it considers neurally plausible (single-valued signals which pass through weighted synapses and are then summed and thresholded), it then condones some transgressions (e.g. error signals in backpropagation) and yet not others (the node and link complexity needed to implement rapid links).

The proposed resolution of this impasse, then, is to cut the Gordian Knot: make a set of simple assumptions that are required to allow elements in the active set to form quick connections to one another *because* they are in the active set and not because they are already directly connected. The one important criterion to apply to whatever mechanism is devised is that it call upon only modest computational resource requirements, such as would be appropriate given connectivity, signal speeds, etc.

The remainder of this section gives a brief description of how one example of a rapid linking mechanism might work, then considers the implications for some related problems.

Rapid links

Suppose that elements are single nodes in a large network, and that there are also many "free" nodes that do not represent anything (they are not yet elements). A typical element will have random, fixed connections to a reasonably large number of these free nodes. Some (probably not all) of the active elements are able to use the free nodes to set up lines to other members of the active set, in much the same way that an analog telephone switching system can form a path between any two points. Several free nodes may need to be included in a given link: they are then to some extent dedicated to that particular association.

Attention

Given the number of elements likely to be active at once, it would not be reasonable to link every one to every other. It is therefore supposed that there is a privileged subset of the active elements that have access to the linkage mechanism: these may well be what define the system's "attentional focus." How is this subset demarcated? This is a subtle question. Part of the answer must be that unusual events in the system can grab the link mechanism. An unusual event would be one which a mutually familiar set of elements could not successfully encode information about the event. In the absence of such "attention-grabbing" events, the attentional focus might often be used to connect either simultaneously occurring elements, or those which are near to one another in a sensory channel.

New Elements

One node in the chain that connects two attentional-focus elements will be used to represent the co-occurrence of the two elements - it therefore becomes a new element in its own right. If there are future occasions on which this new element is activated, it will be strengthened, but if the regularity it represents does not occur again, it will fade, and may eventually be reabsorbed into the pool of free nodes.

Instantiations

Suppose the system looks at an object, so that elements representing its perceptual features become activated. The most abstract of these will be one which represents the identity of the object. This is not isolated, however, because it will be simultaneously active with whatever context is around. Since the exact context and the object are unlikely to have arisen before, it is likely that a new element will be created to represent the co-occurrence. This is an instance node. An instance node is, at first, the encoding of an "episodic memory," but it may then become more "semantic" if the circumstances recur and it is activated again. Usually there will be no exact recurrence and the node will either remain as an episodic trace, or at some later date be re-used.

Since individuals are always represented as instance nodes, multiple instances can in principle be handled. However, what happens to the properties of the superordinate element when an instance arises? For example, suppose there is an element which is a noun-phrase builder, and a sentence is being formulated which needs to contain two noun phrases? Are two instances created with identical properties to the parent, or does the parent in some way "supervise" the behavior of the two daughters?

The obvious thing to do here is to assume, once again, what is wanted: that when an instance node is created, the parent bequeaths its own structure to the daughter element so that the latter can do its own job of handling a set of constraints. The link between the two elements then encodes the fact that they stand in a superordinate-subordinate category relation.

Variables

If an element encodes a pattern of surrounding elements when it is created, it may find that when it is next activated, some of the elements in the original encoding recur, but some do not. If the element is sensitive not just to the identity of its neighbors when activated, but to what *they* are connected to, it may notice that there is a neighbor whose identity is different each time, but which is always connected to some other secondary neighbor which remains fixed. The varying neighbor, then, has a fixed property: it is a variable. Notice that this mechanism requires only that elements can "see" the connections made by a neighboring (connected) element. This is much the same as the assumption made in the previous section, concerning instantiations.

Relationships.

If an element learns to encode a co-occurrence in which two variables take part, it has effectively come to represent a relationship that can hold between members of two classes of object. There is no principled difference between an element that encodes an object and one which encodes relationships between objects, except in the number of variables they tend to accept.

Operations and Analogy

Some elements may encode sequences of events that involve operations on other elements - such an operation is just the internal equivalent of an action on the outside world. But since an operation is itself an element, it too can be operated upon. Finding analogies, or metaphors, has a lot to do with developing ways of modifying operations so that they can be validly performed on other than their original operands. It may well be that the tendency to hunt for analogies is a primitive mode in the system, so useful is it for the creation of novel types of element.

Relaxation

It is perhaps straightforward enough to see how low-level sensory input elements can activate appropriate higher-level elements until eventually a full representation of some perceived object is formed: this is all happening by *relaxation*, because the many elements activated at any point will quickly constrain one another until the set forms a mutually consistent whole, and spuriously activated elements have been suppressed. This relaxation process is maintained, or justified, by incoming sensory signals. Is it possible that the element activations that constitute (i) the motor output representations and (ii) the planning and abstract thought structures, are driven by the same relaxation mechanism? All that relaxation does is to drive all elements to find maximally consistent configurations with respect to one another given certain boundary conditions. In the case of sensory input the boundary conditions are clear, but what is the counterpart for motor structures and planning/thought structures?

It seems necessary to postulate some primitive motivations impinging on the system from, so to speak, below. The planning elements are continually striving to get maximum feedback from something akin to "pleasure" areas, but only certain kinds of activity will cause these to respond (and even then, not for long without some change of scenery). From this point of view, the planning region receives input from the sensory elements and organises appropriate element structures over the motor area which then resolve down into detailed output gestures.

By this means the general concept of relaxation into maximally consistent states (in some cases consistent with the world, other times consistent with something more intangible) can be the driving force for events throughout the system.

Conclusion

It is possible to conceive of a type of cognitive model with basic constituents that are neither symbols nor neuron-like units, but which have characteristics of both. A construct such as this would sit at a level of description higher than the strictly neural level, so it might eventually be realised as either (i) a cluster of neurons, or (ii) an activation pattern across many neurons or (iii) as a single neuron (assuming that there are significant neural properties yet to be discovered). Such a construct might help us to reconcile the otherwise rather disjoint types of explanation provided within the symbolic and connectionist paradigms. However, such systems are complex, and as such are extremely resistant to analytical methods of study. For any empirical investigation to be of value a systematic exploration of these systems would have to be undertaken. This paper has given the briefest of sketches of the type of system involved. Whether they are practically viable and thought worthy of implementation remains to be seen.

References

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-212.

Hinton, G. E. & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum.

Hopfield, J. J. (1982). Neural networks and systems with emergent collective computational properties. *Proceedings of the National academy of Sciences USA, 79,* 2554-2558.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactivation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-407.

McClelland, J. L., Rumelhart, D. E. & Hinton, G. E. (1986). The appeal of parallel distributed processing. In Rumelhart, D. E., McClelland, J. L. & The PDP Research Group (Eds). Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations. Cambridge, Mass.: MIT Press.

Rueckl, J. G., Cave, K. R. and Kosslyn, S. M. (1991). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience, 1*, 2, 171-186.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323,* 533-536.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* (4), 523-568.