Connectionism, Systematicity, and the Frame Problem

W.F.G. HASELAGER and J.F.H. VAN RAPPARD
Dr. W.F.G. Haselager & Prof. Dr. J.F.H. van Rappard Theoretical Psychology Department, Amsterdam / Vrije Universiteit, De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands ⋆

Abstract. This paper investigates connectionism's potential to solve the frame problem. The frame problem arises in the context of modelling the human ability to see the relevant consequences of events in a situation. It has been claimed to be unsolvable for classical cognitive science, but easily manageable for connectionism. We will focus on a representational approach to the frame problem which advocates the use of intrinsic representations. We argue that although connectionism's distributed representations may look promising from this perspective, doubts can be raised about the potential of distributed representations to allow large amounts of complexly structured information to be *adequately* encoded and processed. It is questionable whether connectionist models that are claimed to effectively represent structured information can be scaled up to a realistic extent. We conclude that the frame problem provides a difficulty to connectionism that is no less serious than the obstacle it constitutes for classical cognitive science.

> 'It appeared that Newell and Simon were well on their way to fulfilling the prediction they had made in 1958 that 'in a visible future. . . the range of problems (computers) can handle will be coextensive with the range to which the human mind has been applied.' (. . .) Simon's claims fell into place as just another example of the phenomenon which Y. BarHillel had called the fallacy of the successful first step.' In a talk I gave at RAND, I compared AI to alchemy to make the point. Like the alchemists trying to turn lead into gold, I said, AI had fancy equipment, a few flashy demos, and desperately eager patrons, but they simply had not discovered the right approach to the problem'
>
> Hubert Dreyfus describing his evaluation (originally published in 1965) of Newell and Simon's early work in classical AI (Dreyfus and Dreyfits, 1986, pp.6–7).

Key words: connectionism, distributed representation, frame problem, systematicity

1. Introduction

The frame problem has played a prominent role in debates within cognitive science. It has been claimed to be unsolvable for classical cognitive science (a.o. Dreyfus and Dreyfus, 1987; Horgan and Tienson, 1996) but easily solved by its main competitor connectionism (Churchland, 1989; Meyering, 1993). The frame problem arises when one attempts to model the human ability to keep track of relevant changes in the environment. In general, human beings easily grasp what is going on in their surroundings, as is evident from their capacity to rapidly predict, react or adjust to the important consequences of a certain event. Although many different interpretations of the frame problem exist (Fetzer 1991; Haselager 1997[1]; Hayes

⋆ (e-mail WFG. Haselager@psy. vu.nl, JFH. van.Rappard@psy.vu.nl)

Minds and Machines **8:** 161–179, 1998.

1991; Pylyshyn 1987), the fundamental difficulty, in our view, is that *everything* we know is potentially relevant for our interpretation of what is happening around us. Since we know a great deal, the knowledge we possess must be stored and utilized in such a way that the relevant parts of it are immediately brought to bear on the formation of our beliefs. This imposes heavy demands on both the structuring and the processing of represented information. In this paper we will investigate the potential of connectionism to solve the frame problem. More specifically, we will first of all suggest that the frame problem can be regarded as raising questions concerning the adequacy of the classical, symbolic, representational format. We will indicate why from this, what we will call, 'representational' perspective distributed representations may be regarded as promising with respect to the frame problem.

Secondly, we will claim that since the frame problem arises in contexts where a potentially large amount of often complex knowledge is involved, connectionism has to prove that its models can represent and use well-structured information. Since this matter has been extensively discussed in relation to the issue of systematicity, we will, thirdly, examine some proposed connectionist solutions to the problem of systematicity that are generally regarded as promising. We will argue that such models are as of yet unsatisfactory and moreover unlikely to be scaled up successfully to more realistic, complex tasks. In all, we argue that although the difficulties encountered by connectionism when addressing the frame problem may be of a different kind compared to those of classical cognitive science, they are no less serious.

2. Connectionism and the Representational Approach to the Frame Problem

Among the many problems cognitive science encounters an interesting and hotly debated one is the so-called 'frame problem'. In the history of AI, the frame problem was first encountered (and named) by McCarthy and Hayes (1969) in their attempt to create a general intelligence on the basis of a strictly deductive inference mechanism. Their model decided what to do by deductively inferring that a certain sequence of actions or events would lead to a desired goal. An unfortunate consequence of this strategy, however, was that the model would need not only rules specifying what would change because of an event but also rules indicating what would remain the same. Otherwise the model would not be able to deduce the new situation. Because of the overwhelming amount of rules specifying non-changes the system would simply get lost in performing irrelevant deductions. As such the frame problem has played an important role in many developments in logic (c.f. Shanahan 1997). Since the article by McCarthy and Hayes, the frame problem has also become known as a more general difficulty for cognitive science. This has led to sometimes chaotic discussions, as there seems to be little agreement on what exactly the frame problem is, what the main reasons of its emergence are, how it should be solved, and what would count as a solution[2] (c.f. Haselager, 1997). In the context of this paper, we will not enter this debate, but simply focus on the frame problem in the more general sense as an obstacle encountered when trying to understand the psychological mechanisms involved in common-sense reasoning, instead of as an issue in logic.

Psychologically speaking, people have an amazing ability to quickly *see* the *relevant* consequences of certain changes in a situation. They *understand* what is going on and are able to draw the *right* conclusions quickly, even if this means retracting earlier beliefs and adopting new ones. The problem is how to model this ability computationally. What are the computational mechanisms that enable people to make common-sense inferences? Especially, how can a computational model be prevented from fruitlessly engaging in time-consuming, irrelevant inferences? A rather straightforward suggestion is that seeing the relevant consequences of an event is made possible by an understanding of the situation. One reaches an understanding of the situation by *using what one knows*. Yet, human beings posses an enormous amount of information. The real difficulty underlying the frame problem is how the *relevant* pieces of knowledge are found and how they influence one's understanding of the situation. Regarding this issue, it seems to us that two general strategies can be discerned.

First of all, one can see the frame problem as the problem of how, on the basis of what one knows and perceives, one can quickly generate a plausible interpretation or hypothesis to explain what is going on. This leads to an examination of the problems concerning non-demonstrative inference (i.e. induction, abduction, and inference to the best explanation). This *inferential* approach focuses on the reasoning process itself and investigates how more global characteristics of what one knows can guide the search for plausible inferences (c.f. Haselager, 1997). Secondly, one can see the frame problem as something to be solved by the development of a good representational format. This *representational* approach is characterized by a strong interest in questions of what kind of information should be regarded as primary and what kind of representation is most suited for representing this information in such a way that the appropriate inferences will ensue almost automatically. In the following, we will concentrate on the representational approach to the frame problem and its relation to connectionism.

Janlert (1987) does not make a distinction between an inferential and a representational approach to the frame problem but, in our view, he provides a good illustration of the representationalist perspective. The main issue here is how to adequately represent a changing world in a computationally *efficient* manner (Janlert, 1987, p.7–8). A decision on what has to be represented has to be supplemented by a decision on how to represent and process it. Janlert (1987, pp.2, 37–38) suggests that information might have to be represented *intrinsically* in the system. The concept 'intrinsicness' should be understood as stressing that any approach to the frame problem needs to refrain from completely describing everything there is to know about the world: it should avoid having to 'spell it all out'. Janlert refers to Palmer (1978) according to whom a representation is intrinsic if a representation of a relation has the same inherent constraints as the relation itself (Palmer, 1978, pp.271–272). That is, the constraints are not arbitrary and not imposed from outside (i.e. do not have to be stated explicitly) but follow from the inherent structure of the representation. Haugeland (1987) gives the example of a scale model that has a structure similar to the domain it models (the main difference being, of course, the scale). Therefore, every consequence of an event in the actual world will be precisely matched by the consequences of the small-scale event in the small-scale representational world. The result is, as Haugeland (1987, p.86) puts it, that representations of side effects of events are just side effects of the representations of those events.

Perceiving the issue in this way, the main source of the problem for classical cognitive science is that propositions and their symbolic counterparts do not represent properties and relations intrinsically (e.g. Palmer, 1978, p.296). According to Haugeland (1987), the propositional symbolic representational scheme harbors a distinction between what is *explicitly* and *implicitly* represented. Explicit representations being directly usable, whereas implicit information first has to be explicitly inferred (i.e. logically derived and symbolically represented) before it can be used (Haugeland 1987, p.90; 1991, pp.74–75). Therefore, in a symbolic representational system, *all* the objects, constraints and relations found to obtain in the world have to be specified *explicitly*. This is problematic for if everything is explicitly represented there will be great problems in quickly locating a particular represented item (Janlert, 1987, p.36). The use of a representational system that enforces an explicit representation of all information before it can be used by the system may therefore be considered as the source of the frame problem. A better approach might be to use intrinsic representations for which a distinction between what is explicit and what is implicit cannot even be made (Haugeland, 1987, pp. 88–91).

From the representational perspective, then, the frame problem demonstrates the need to find an intrinsic way of representing information. An important point proved to be the fact that no distinction between explicitly and implicitly represented information should be allowed. On the basis of this analysis of the frame problem, we suggest that connectionist models using distributed representations are valuable candidates for further inspection.

3. Distributed Representations and Prototype Activation

In the following we assume that the basic architecture and standard interpretation of connectionist models is familiar (c.f. Bechtel and Abrahamsen, 1991; Churchland and Sejnowski, 1992; Churchland 1995). Cognition is understood in terms of the transformation of patterns of activation. Weights attached to the connections between the units of the network collectively determine the nature of the activation pattern transformation (Churchland, 1989, pp. 201–202). Weights represent the enduring knowledge of the network and determine how the network will react to incoming stimuli. A specific set of weights, as embodied by a trained network, is a prime example of *distributed representation*. The notion of distributed representation is a fundamental one in connectionism, yet its precise meaning long remained obscure[3]. Van Gelder (1991b) has written a survey of the concept of 'distribution' as it occurs in the literature. He concludes that the notion of super(im)position of representings[4] over a portion of representational resources is the most common theme in discussions on the nature of distributed representations (Van Gelder 1991b, p.42). A representation is distributed if it is representing many items while using exactly the same resources (Van Gelder 1992, p.176). Importantly, there is no direct relation between a single weight and a single represented item, but instead all weights partake in representing all information the network possesses. The representation of distinct items is superimposed on the same set of representational resources (Van Gelder 1991b, p.43). No part of the representation should by itself be able to represent a distinct content. No matter how the representational resources are sliced, each content item must be represented over the same extent of the resources as the others (Van Gelder 1992, p.178). To put things differently; the representings of distinct items are superposed if they occupy the same set of representational resources (Van Gelder 1991b, p.43). The greater the total content that is represented by the same amount of representational resources, the more superposed the representing is (Van Gelder 1991b, p.44), and the more distributed the representation. Distributed representations, then, are essentially characterized by the superposition of representings.

Van Gelder (1991b, p.55) notes that distributed representation (as characterized above) is deeply affiliated with the connectionist approach in that neural networks provide a natural medium for implementing them. Furthermore, distributed representations are theoretically significant because they are radically different from the classic symbolic representational approaches where specific representations correlate with specific represented elements (Van Gelder 1991a, p.373).

Semantically, the configuration of connection weights embodies the knowledge the network possesses, yet it cannot in any straightforward propositional sense be semantically interpreted. In fact, since a specific configuration of weights determines every reaction to every input the network is capable of, its meaning would be the total of all the potential reactions to incoming stimuli which the network is capable of, i.e. the sum of all interpretations of all possible activation pattern transformations taken together. Van Gelder (1991b, p.54, see also pp.34, 45 and p. 55), therefore, rightly speaks of a 'fundamental gulf in kinds of representation'. This fundamental difference between distributed and symbolic representations reflects a basic disagreement about the nature of the states that are taken to encode the information an organism possesses, and the way this information can be utilized in the production of behavior.

From a connectionist perspective, an individual's knowledge of his or her surroundings consists of a set of weights, or alternatively, a point in weight space. The function of the weights can be thought of as a *partitioning* of the activation space of the hidden units. During the learning phase, the weights of a network are set in a way that results in a useful partitioning of the activation space of the hidden units. The clearest cases learned ('prototypes') will occupy a specific region inside a partitioning (a so-called 'hot spot'), the least clear cases will occupy places near the border between partitionings. A prototype can be thought of as a point or small volume in the hidden unit activation space (a 'hot spot'), representing a family of relevant features that are characteristic for stimuli belonging to a specific kind (Churchland, 1989, p.206). Different input vectors can result in the activation of this prototype vector by the hidden units: the network has learned that these diverse inputs are similar with respect to the task it has learned. Previously unencountered input patterns that result in hidden unit activation patterns close to the prototypical region will evoke the prototypical reaction of the network. Activation patterns that occupy a place at a considerable distance from the prototypical region in the activation space indicate that the system is dealing with a murky case. The network will react most unambiguously to prototype cases. The knowledge possessed by an organism consists mainly of a substantial set of prototypes:

> 'The picture I am trying to evoke, of the cognitive lives of simple creatures, ascribes to them an organized 'library' of internal representations of various prototypical perceptual situations, situations to which prototypical behaviors are the computed output of the well trained network.' (Churchland, 1989, p.207; see also 1995, p.83).

Churchland (1989, pp.212–218; 1995, pp.97–143) suggests that the model applies to a wide range of prototypes (a.o. categorical, temporal, social and motivational) and thereby provides a unified account of much of our explanatory understanding.

According to Churchland, it is on the basis of this connectionist model that we can understand how human beings can see the relevant consequences of events or actions in their environment so quickly. Explanatory understanding consists in the activation of a prototype vector (Churchland 1989, p.210, see also p.208). The network generates an explanatory hypothesis, which effectively says: 'this incoming information is of such and such a type'. The generation process itself, though fast, can be quite complex. Since a network may have many layers of hidden units, there is ample possibility for quite complex processing. Furthermore, a network can receive input from other areas of the brain and, through recurrent connections, transmit the results of previous processing back into its earlier layers (Churchland 1989, pp.208; 1995, pp.99–108). Churchland also indicates that the prototype activation model does not instantiate a 'mere' process of classification but adds information to the incoming activation pattern. The activation of a prototype models explanatory understanding as a kind of *ampliative recognition*. An organism ends up understanding far more about the situation than was originally present in its input since the prototype is the result of the previous complex processing of many examples during its learning stage (Churchland 1989, p.212).

We think that the essence of a connectionist solution to the frame problem is clear. Cognitive systems can quickly see what the relevant consequences of a change in the environment are because the information they receive results in an almost instantaneous activation of an adequate prototype that constitutes an explanatory understanding of the situation. As all weights influence the emergence of the activation vector, everything the network knows is brought to bear upon its response to incoming information. Churchland (1989, p.178) argues that because the relevant knowledge is activated automatically, the frame problem does not even arise. He is supported in this optimistic attitude by several others. Meyering (1993, p.31), for instance, has expressed great confidence: ‘the infamous frame problem is solved in the blink of an eye’ (our translation). It bears emphasis that this proposed solution to the frame problem is made possible precisely because connectionism forswears the symbolic representational format. The use of symbolic representations as utilized in classical cognitive science allows for a close mirroring of the structure of knowledge as propositionally described. The relationships between concepts as expressed in a proposition are explicitly represented by links between symbolic structures in a hierarchy or through the use of rules. To capture the potential relevance of everything to everything, every possible relation between two concepts which might at a certain point in time and in a certain context become important needs to be represented explicitly through a hierarchical link or a rule. Even if this were feasible, finding the relevant information in the midst of a myriad of symbolic structures and their interconnections quickly becomes computationally overwhelmingly complex.

Modeling information processing on the basis of distributed representations and activation pattern transformation seems to skip these problems because no attempt is made to match closely a propositional specification of information. Instead, a distributed representation represents information in holistic fashion, without its decomposition into constituent concepts and their interrelationships. As the knowledge of a system lies embodied in its weights, it directly and automatically constrains the processing of incoming information. There is no need to search for the relevant pieces of information before they can be applied. Moreover, changing the knowledge of a system after an event has occurred need no longer take the form of an explicit reconsideration of all symbolic structures and their interconnections. Changing the setting of one weight automatically influences all the information processing the network is capable of.

In the terminology of the representational approach to the frame problem, distributed representations may be said to represent their information intrinsically. In our view, distributed representations do not allow a distinction between explicitly and implicitly represented information. Claiming that distributed representations represent their informational content explicitly seems rather awkward for, because of the superposing of representings, it cannot even be clearly stated what they would explicitly represent (other than: everything the network knows). They also do not represent implicitly because nothing has to be derived before it can be used for processing. Instead, the weights directly and immediately influence the processing of information. It is interesting to note that the difficulty of applying these notions to connectionist ways of representing information has led to a reconsideration of the nature of the explicit-implicit distinction (Clark 1992; Hadley 1995; Kirsh 1990). Rather than drawing the conclusion that these notions are inapplicable, however, attempts have been made to redefine them in various ways. This has led to rather counterintuitive results (e.g. Kirsh's claim that symbolic representations need not be explicit) and contradictory conclusions about distributed representations being explicit (Clark) or implicit (Hadley). Like Elman (1991, pp.218–219) we doubt the usefulness of the distinction with respect to distributed representations and we agree with his claim that it makes no sense to view networks 'through traditional lenses'. In our view, then, distributed representations conform to the requirements of the representational approach. They do not allow a clear distinction between explicitly and implicitly represented information but represent their informational content intrinsically. So far, it seems clear that distributed representations can be regarded as most promising with respect to the frame problem. However, to see whether the frame problem really is that easily solved, we propose to look at an issue that lies beneath the surface of the suggestion of distributed representations.

4. The Representational Capacities of Distributed Representations

Connectionist research normally focuses on relatively small networks attempting to solve restricted tasks. But one of the characteristics of the frame problem (as classical cognitive science belatedly discovered) is that it shows up especially in more realistically complex situations. The question therefore is whether the basic suggestions examined above can be easily 'scaled up'. How can connectionist models handle large amounts of knowledge? Are distributed representations really adequate when it comes to the representation of complex information involved in reasoning and understanding? Or is the gain in automatic and direct retrieval or resonation of relevant knowledge overshadowed by a substantial loss in the capacity to represent the structure of information? Serious doubts have been ventured in this respect (e.g. Fodor and Pylyshyn 1988; Holyoak, 1991, pp.315–316; Thagard 1992, pp.242–243).

As is well known, the classical solution to the problem of representing structure invokes the use of a representational format with a *concatenative constituent* structure, resulting from the part/whole relationship between simple and complex representations. The simple elements out of which complex representations are construed are literally present in the complex representation. In a symbolic representational system, the concatenative constituent structure of complex representations is utilized to match the structure of the information represented. The relations between elementary representations explicitly represent the relations between parts of the information (e.g. that in 'John loves Mary' it is Mary being loved, not John). Fodor (1975; Fodor and Pylyshyn 1988) has argued that in order to explain certain characteristics of cognition that are referred to by the terms 'productivity[5]' and 'systematicity[6]' it is necessary for the representational system to be compositional, meaning that representations have a combinatorial syntax and semantics, which is made possible by their concatenative constituent structure. Although the exact nature and pervasiveness of productivity and systematicity are open to discussion, it is generally agreed that a system must be able to represent complex structured information in order to exhibit interesting cognitive functions. Any representational scheme that is of interest to cognitive science must, at least to a considerable extent, be compositional (Chalmers 1993, p.306; Hinton 1990, pp.2–3; Pollack 1990, p.78; Smolensky, 1991, p.288; Van Gelder, 1990, pp.355-356).

Van Gelder (1990; 1991a) and, more recently, Horgan and Tienson (1996) have pointed out that the part/whole relation between complex representations and their constituents is not the only formal relation available for the encoding of causally effective constituent structure. Horgan and Tienson state, correctly in our view, that

> 'the question is not whether *constituents* can play a causal role. The question is whether the fact that a representation *has* a particular constituent can play a causal role. And that fact can play a causal role *if* the representation carries the information that it has that constituent.' (p.79, see also p.74).

In other words, the constituent need not be physically present as long as the information it carries is effectively manifested in the encompassing representation (p.80). Syntax does not entail a part/whole relationship, but merely the systematic and productive encoding of semantic relationships (p.71, 73).

Horgan and Tienson claim that there are examples of a rudimentary kind of structure-sensitive processing of non-classically structured representations. They are, furthermore, quite optimistic with respect to the potential of connectionist models, regarded as a dynamical systems, to preserve structure in the representations. Indeed, they assert that relations between strategically positioned representational points in a properly molded activation landscape can embody structural relations as rich as and even richer than the classical symbols-rules representational scheme:

> 'The structural resources are certainly there, much more so than in classicism: high-dimensional dynamical systems can have structure far richer than the intrinsic structure of computing machines, and positional relations among points in a dynamical system can exhibit structure far richer than the intrinsic structure of classicist representations.'(p.163, see also p.154).

They refer to work by a.o. Pollack and Chalmers as examples of the kind of non-classical syntax they endorse. However, in our opinion it is far from clear whether the models proposed actually succeed in displaying systematicity. Furthermore, we will indicate reasons to doubt that the mechanisms used allow a scaling up to realistically complex contexts.

As Van Gelder has put it, the challenge to connectionism is

> 'to devise models in which structure-sensitive processes operate on the compound representations themselves without first stopping to extract the basic constituents. These processes must capitalize directly on the inherent and systematic similarities among the nonconcatenative representations, (Van Gelder 1990, p.381; see also Chalmers 1993, p.312; and Fodor and McLaughlin 1990, p.202, no. 14).

David Chalmers (1990; 1993) has attempted to meet this challenge. Chalmers presents a connectionist network utilizing distributed representations that models the transformation of sentences in the active to the passive mode. He uses syntactic transformation as an example of structure-sensitive operations (Chalmers 1993, p.313). For instance, the sentence ‘John loves Michael’ should be transformed by the network into ‘Michael is loved by John’. Note that the information present in the structure of the sentence is that John is the one that loves Michael, and not vice versa. On Fodor’s account, a network is incapable of distinguishing between ‘John loves Michael’ and ‘Michael loves John’ since it is ‘structurally blind’ and merely associates ‘John’ ‘loves’ and ‘Michael’. Providing a transformation into the correct passive mode, then, indicates that the network is able to recognize and use the structure in the information represented.

First, syntactically structured sentences (represented by trees) are transformed into distributed representations. This is accomplished by Pollack’s (1990) RAAM network[7]. The resulting distributed representations are used by the actual transformation network (a basic, three layer feedforward network, learning through back-propagation) that performs the passivisation directly on the distributed representations without using a decomposition process first. The resulting output is of course again a distributed representation which is then fed into the RAAM, translating it back to its syntactic structure. The question, of course, is whether Chalmers’ network is able to use the structure that is implicitly contained in the activation patterns provided by the RAAM network.

After training, the transformation net was tested with new sentences. Chalmers (1990, p.60) reports a 65% generalization rate, which, high in itself, went up to 100% after correction of RAAM errors. Chalmers concludes:

> ‘Not only is compositional structure *encoded* implicitly in a pattern of activation, but this implicit structure can be *utilized* by the familiar connectionist devices of feed-forward/backpropagation in a meaningful way’. (Chalmers 1990, p.60; see also Chalmers 1993 p.314).

So, Chalmers claims, his results contradict Fodor’s thesis that concatenative constituent structure has to be present in representations in order to be of use to information processing mechanisms. There is no need for an explicit tokening of the simple parts of the representation in the complex one. Distributed representations can have enough formal structure to be functionally compositional and of direct use to the system’s processing.

5. Lawfulness versus Coincidence

Importantly, Horgan and Tienson (1996, p.80) claim that the existence and useability of nonclassical representations that carry constituency information can be read off of systems like Chalmers’ that perform constituent-sensitive operations. That is, it is the *performance* of the models on which their claim of non-classical effective syntax is based:

> ‘It is quite clear that tensor-product representations and RAAM representations do carry constituency information within or relative to a system, and that this information is available to the system. It is clear because the systems perform constituent-sensitive operations. That the representations carry this information is shown by the whole system of dispositions of the successful system’. (p.80).

They stress that it is the capacity of the system to 'perform properly on inputs not among the training corpus.' (p.75) that substantiates claims for a rudimentary form of effective syntax.

Yet, precisely with respect to the network's capacity to generalize to novel input serious criticisms can been raised. For instance, Hadley (1994a, p.261) notes that the novel corpus of sentences that Chalmers used to test his network contained no new words (i.e no words not already encountered during training) nor words occupying new syntactic positions (i.e the network had encountered all words in all syntactically possible places during training). In other words, the novelty of Chalmers' corpus of test sentences is rather moderate[8]. As Hadley says (1994a, p.262), if a completely new word were introduced in an otherwise familiar sentence, this might result in such disruption of the network that it would not even recognize the familiar lexical items.

Hadley concludes that Chalmers' model does not succeed in capturing the kind of systematicity argued by Fodor as being characteristic of human cognition. Concerning this problem, we think that Hadley's proposed criterion of generalization ability, i.e the network's capacity to deal with genuinely novel sentences, is adequate. Hadley (1994a, p.271) notes that in the light of this criterion the work of Chalmers and several other connectionist attempts (including the work of Pollack, Smolensky and others) to answer Fodor's challenge do not succeed in displaying the strong degree of systematicity characteristic of humans.

Hadley's criticism raises the important and more general point that one has to be very careful that the structure sensitive behavior of a network is not simply the result of prearranged statistically large similarities between the training data to which the network has become tuned and the test data. This hampers a straightforward assessment of the force of connectionist examples of structure sensitive processing.

The matter of distinguishing real systematicity from prearranged statistical coincidence also comes to the fore in the discussion about Fodor's repeated claim that merely providing counterexamples is far from sufficient to show that connectionism can deal with compositionality in a completely satisfactory way. As he says, it is a *law* that cognitive capacities are systematic (Fodor and McLaughlin, 1990, pp.202–203; Fodor and Pylyshyn 1988, p.48). That is, it is easy to 'wire up' a non-systematic connectionist network, but it is impossible to create an unsystematic classical system. The point of the law-requirement, as Butler (1993, p.323) notes, is that merely showing that systematicity is *possible* on the basis of a connectionist architecture is not enough; it must be indicated why systematicity is *necessary* given the architecture. Likewise, Butler continues, a theory of planetary motion that merely allowed for the possibility of elliptical orbits of planets would be considered as insufficient. To really count as an explanation, it would have to show that the nature of such orbits necessarily followed from the theory. Similarly, connectionists have to demonstrate that systematicity necessarily follows from the architecture[9].

6. The Plausibility of Learning Conditions

Now, one can, as does Chalmers (1993, p. 316), quite rightly point out that the fact that 'differently wired' networks could easily be insufficient at best shows that not *all* possible connectionist architectures are satisfactory. To be acceptable, Chalmers says, the class of rightly wired networks would have to be compositional and display systematicity under many different learning conditions. We think Chalmers is right in this, but it only helps to underscore the fact that merely demonstrating that a rightly wired connectionist network with distributed representations can be compositional is not sufficient, since this might be an artificial result of the specific characteristics of the training and testing data. Fodor's requirement that systems *must* be compositional can, we suggest, most beneficially be seen as an attempt to provide a safeguard against too readily taking 'accidental' signs of systematicity for the real thing. We propose, then, to take the requirement of displaying systematicity under many different (or at least psychologically realistic) learning conditions as a second constraint, in addition to the generalization requirement discussed above.

The importance of this second constraint becomes clear if one considers connectionist attempts to deal with Hadley's generalization criterion. For instance, Christiansen and Chater (1994) present two simulations, one in which the network failed to exhibit strong generalization (in a genitive context) and one in which it succeeded (in the context of noun phrase conjunctions). e.g., when presented with the sentence 'Mary's girls run' (where 'girls' had never occurred in a genitive context in the training set), the network failed to behave similarly to 'Mary's cats run' ('cats' having occurred in the genitive context in the training set). However, when presented with 'Mary says that John and boy from town eat' ('boy' not occurring in a noun phrase conjunction in the training set), the network correctly predicted a plural verb, thereby making the strong generalization that a noun phrase conjunction (even an unfamiliar one) requires a plural verb.

Although Christiansen and Chater (1994, p.285) conclude on the basis of their work that future progress is possible, in our view these mixed results underline the importance of Fodor's law requirement. Why did the network succeed in the context of noun phrase conjunctions but fail in the genitive context? Christiansen and Chater do not present a principled explanation of these results. In our view, this considerably detracts from the value of their models. After all, one would like an explanation of systematicity, not just a mere demonstration (see also Niklasson and Van Gelder, 1994, p.297). Furthermore, Christiansen and Chater suggest that the network might be able to succeed 'if a different kind of representation is used or the details of the training are altered' (1994, p.282). But it is exactly this kind of 'fetching' that Fodor's law-requirement is aimed at preventing.

A second point of concern involves the enormous amount of training that is necessary before the network can be said to have learned its task. Christiansen and Chater (1994, p.280) report a total of 32 epochs, each one presenting the network with the full training corpus of 10.000 sentences for a relatively simple phrase structure grammar (6 rules) and a small vocabulary (34 items[10]). It seems unavoidable that the amount of training needed will become unmanageable in the case of a grammar and vocabulary of a realistically large size. Finally one has to notice the complexity of the training setup, with many carefully arranged details (including, for instance, the periodic resetting of context units[11]). Given the constraint, presented above, that systematicity has to be demonstrated under a variety of learning conditions, this provides a further reason to regard the results as unconvincing.

A second connectionist attempt to answer Hadley's generalization constraint, by Niklasson and Van Gelder (1994), concentrates on the case where test sentences contain at least one atomic constituent that did not appear anywhere in the training set. Using the same kind of architecture as Chalmers, they introduce a novel symbol ('s') to the network that has been trained to transform formulas according to the following inference rule: $p \rightarrow q \Leftrightarrow, \neg p \vee q$. The network succeeds in handling formulae containing the new symbol after a huge amount of training (4000 passes through the training set of 600 formulae (p. 294–295; they speak of an 'exhaustive exposure to a training set', p.299). Niklasson and Van Gelder (1994, p.298) conclude that points in an activation space can function as representations in a way that allows spatial structure to preserve syntactic structure useable for further processing. However, we want to emphasize that a proper localization of representational points within the spatial structure has been *prearranged* by Niklasson and Van Gelder by means of a seperate RAAM network, called the 'representation generator[12]. As they say:

> 'The design and training regime of the representation generator results in representations that are systematically positioned in the space so that the representation for 's' occupies the space in between the 'known' constituents' (Niklasson and Van Gelder, 1994 pp.297–298; our emphasis; see also Hadley, 1994b, pp.437–438).

As in the case of Christiansen and Chater, we conclude that the results are largely dependent on a meticulously designed architecture and training regime, thus violating the constraint that systematicity has to be demonstrated under a variety of learning circumstances.

In all, we conclude that connectionist models as presented by Chalmers, Christiansen and Chater and Niklasson and Van Gelder depend for their limited successes on very strict, carefully arranged and psychologically unrealistic learning circumstances (i.e. the amount and details of training). Hence, we do not think there are good reasons to expect that models of this kind will succeed when confronted with more realistically complex tasks. Yet, it is precisely under these more realistic circumstances that the frame problem arises, so we fail to see how connectionism would be able to deal with that problem successfully. Before drawing our final conclusions, we will point briefly to a further difficulty, in addition to the problems of generalization and the specificity of learning conditions, that may make the idea of functional compositionality seem even less promising.

7. Interacting Distributed Representations

Granting for the moment that structured information might be adequately represented by distributed means under a variety of learning conditions, there is the further issue of how representations of this kind can *interact*. This is of especial relevance to the domain of commonsense reasoning, the area in which the frame problem looms large. In this respect, it is remarkable that connectionist attempts to model common-sense reasoning ultimately refrain from using *fully* distributed representations and instead use a hybrid (if not completely classical) representational format, as (each to a different extent) in the case of Derthick (1990); Miikkulainen and Dyer (1991); Shastri & Ajjanagadde (1993) and Sun (1994).

For instance, a recent and ingenious connectionist model of common-sense reasoning is outlined by Shastri and Ajjanagadde (1993). The representational theory that they implement in a connectionist architecture is a rather classical one of (complex) facts, rules and conceptual hierarchies. Moreover, they explicitly reject the use of distributed representations as being unsuited for representing large amounts of structured knowledge, because it 'cannot have the necessary combination of expressiveness, inferential adequacy and scalability' (p.485). The problem is that when distributed representations are combined into more complex (and still distributed) representations, a loss of binding information (e.g. as to which objects are bound to which predicates) seems unavoidable. Hummel and Holyoak (1993) similarly point out that distributed representations of, for instance, predicates and objects cannot be combined into larger distributed structures, without losing information about which objects are bound to which predicates. For instance, a distributed representation of 'Ted gave Mary flowers' is difficult to combine with a distributed representation of 'Jane knows that p' into a distributed representation of 'Jane knows that Ted gave Mary flowers' without losing information as to who knows what or who gave what to whom. That is, there is an 'inherent tradeoff between distributed representations and systematic bindings among units of knowledge' (p. 464) that becomes clear as soon as several distributed representations have to be combined. We want to emphasize that the difference with the task RAAM is fulfilling is that in the case of RAAM *non*-distributedly represented constituents or complexes are added to a distributed representation, whereas in the case discussed by Hummel and Holyoak distributed representations are added to *distributed* representations.

So, even if one assumes that proposals a la Chalmers, Christiansen and Chater or Niklasson and Van Gelder ultimately might satisfy the constraints of generalizing under a variety of learning conditions, the problem is not completely solved. Even if one may accomplish structure sensitive processing of distributed representations of chunks of information *separately*, the applicability of such a proposal to realistically complex cases, where distributed representations of structured information have to *interact* in various ways, remains blocked.

8. Conclusion

The frame problem is generally regarded as a serious, sometimes even unsolvable, difficulty for classical cognitive science. According to the representational approach, the fact that the symbolic representational format allows for a distinction between what is explicitly and implicitly represented is the underlying cause of the problem. Distributed representations can be seen as promising precisely because of their intrinsic nature. However, although we consider the potential of distributed representations to be most interesting, we do not think that an easy victory awaits connectionism. Since the frame problem involves the use of a substantial amount of interrelated knowledge, representing the structure of information is an essential precondition for making progress. We have analyzed connectionist attempts to represent and utilize structure in relation to the issue of systematicity. Although the results are generally presented and regarded as favorable to connectionism, we have indicated three reasons for a more negative appraisal. Genuine doubts may be raised about whether the performance achieved indicates the true capabilities of distributed representations or whether they largely depend on the specifics of the training and testing data. The capacity to generalize is small, and has not been demonstrated under a variety of learning circumstances. Moreover, the capacity of distributed representations to preserve the structure of the information while interacting with other distributed representations seems severely limited. The conclusion must be that connectionism has no principled and satisfactory way of effectively representing structured information in a distributed way. Even if distributed representations could be shown to be successful on small-scale problems (of the kind investigated by Chalmers and others), it is hard to see how their range of application could be extended to a more serious level of complexity. This, in turn, implies that connectionism still has to prove that its models are able to deal with realistically complex situations and events, as classical cognitive science is still trying to do. The connectionist approach to the frame problem may have, in comparison with the classical approach, different problems to cope with, but these present no less significant obstacles to overcome. Like classical cognitive science, connectionism too will have to overcome the fallacy of the successful first step.

Notes

1. This paper uses some material taken from Haselager (1997). Permission by the publisher is gratefully acknowledged.
2. To indicate this, it should suffice to say that the question has been raised whether or not the frame problem can correctly be interpreted as (being related to), in alphabetical order: the bookkeeping problem, the extended prediction problem, the inertia problem, the problem of the metaphysical adequacy of representations, the problem of non-demonstrative inference, the problem of ordinariness, the problem of persistence, the prediction problem, the qualification problem, the ramification problem, the truth-maintenance problem, and the updating problem.
3. Feldman says about the notion of ‘distributed representation’:

> 'The problem is that people have been using this term to denote everything from a fully holographic model to one where two units help code a concept; thus, the term has lost its usefulness.'

(Feldman 1989, p.72; see also Van Gelder, 1991b, p.35).

4. Van Gelder refrains from speaking of superposed representations, and prefers the phrase superposed representings' to stress that there is only one single representational item representing all content (Van Gelder 1991b, p.43).
5. *Productivity* refers to the thesis that, in principle, a cognitive system can entertain an infinite number of thoughts. This indicates that the representational capacities of a cognitive system are, in principle, unbounded. The only way to achieve this by finite means, Fodor argues, is through a representational system that has a combinatorial syntax and semantics (Fodor 1975, p.31–32; 1987, p.137, pp.147–148; Fodor and Pylyshyn, 1988, pp.33–37).
6. The term '*systematicity*' refers to the fact that the ability to understand and/or produce certain thoughts is intrinsically related to the ability to think other thoughts. If a person is capable of entertaining a thought like 'John loves the girl', he or she is bound to be able to have the thought 'The girl loves John' as well. This can be explained by means of the compositionality principle in the following way. The elementary mental representations (atoms) that together represent the content of the thought have a structured relationship (e.g. subject – predicate – object) to one another. The structural relations are the same in both thoughts, only certain atoms have changed place. Understanding or entertaining the first thought means that both the atoms and the structural relations are understood, hence the other thought must be understood as well.
7. RAAM is an acronym for 'recursive auto-associative memory'. Basically, it is capable of representing the information inherent in symbolic tree structures of arbitrary depth as distributed activation patterns. It can compress the representations of the terminal nodes into one activation pattern which represents their parent, and then, recursively, compress all parents one layer up into another single pattern, compress these patterns yet again, etc., thus working from the leaves to the root. Similarly it can reconstruct the children from the distributed representation of the root reconstructing them recursively until the leaves are reached (Pollack 1990, p.84). The RAAM-architecture is general (it applies to tree structures of arbitrary depth), effective (the (de)composing processes are performed mechanically) and reliable (after sufficient training) (Chalmers 1990, pp. 55–56).
8. According to Hadley (1994a, pp. 250–251), a network exhibits weak systematicity if it can handle test sentences that contain words that occur only at syntactic positions already occupied by these words in the training set. The training set is then fully representative of the test set. A system exhibits strong systematicity if it can exhibit weak systematicity and moreover can process novel simple and novel embedded sentences containing familiar words in new syntactic positions. Hadley (pp. 252–254) points to much empirical evidence that children exhibit systematicity in this strong sense.
9. Of course, one may question whether it is really a law of nature that you can't think aRb if you can't think bRa' as Fodor claims (Fodor and McLaughlin, 1990, p.203). The 'lawfulness' of systematicity has indeed been doubted by several writers (Dennett 1991, p.27; McNamara 1993, p.114; Wilks 1990, p.331). However, though there may be room for discussion about the exact extent of systematicity, it is quite clear that children and adults display systematicity in a strong sense (Hadley, 1994a, pp.252–254, p.270).
10. Christiansen and Chater (1994, p.279) specify that the vocabulary consists of 2 proper nouns, 3 singular nouns, 5 plural nouns, 8 verbs in plural and singular form, a singular and a plural genitive marker, 3 prepositions and 3 nouns indicating locations
11. The point we are making is not that no constraints on the training setup are allowable. Rather, the details of training should be reasonably general and justifiable on psychological grounds. For instance, the periodic resetting of context units as used by Christiansen and Chater might not be completely devoid of psychological plausibility. Elman (1993) and Clark and Thornton (in press) argue that a periodic resetting of context-units provides for a kind of limited memory that allows the system to learn the most basic distinctions first. In later phases, the window can be enlarged (by resetting the context units after longer intervals), so the network can learn the finer distinctions, necessary to fulfil its task. They refer to psychological evidence that developmental

limitations of this kind exist and are beneficial. This aspect of the model would therefore satisfy the second constraint. Our point is that most details of the training regimes lack such justification.

12. The representation generator creates distributed representations for atomic constituents, by encoding the tree structures containing type information about the constituents, e.g. whether they are connectives, propositions or symbols (Niklasson & Van Gelder 1994, pp.296–297).

References

Bechtel, W., and Abrahamsen, A. (1991), *Connectionism and the Mind*, Oxford: Blackwell.

Butler, K. (1993), Connectionism, Classical Cognitivism and the Relation between Cognitive and implementational Levels of Analysis, *Philosophical Psychology*, 6(3), pp 321–333.

Chalmers, D.J. (1990), 'Syntactic Transformations on Distributed Representation, *Connection Science*, 2(1 & 2), 53–62.

Chalmers, D.J. (1993), 'Connectionism and Compositionality: why Fodor and Pylyshyn were Wrong, *Philosophical Psychology*, 6(3), pp.305–319.

Christiansen, M.H., and Chater, N. (1994), Generalization and Connectionist Language Learning, *Mind & Language*, 9(3), pp.273–287.

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

Churchland, P.M. (1995), *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*, Cambridge: MIT Press.

Churchland, P.S., and Sejnowski, T.J. (1992), *The Computational Brain*, Cambridge: MIT Press.

Clark, A. (1992), 'The presence of a Symbol', *Connection Science*, 4, 3 and 4, pp. 193–205.

Clark, A., and Thornton, C. (*in press*), 'Trading Spaces: Computation, Representation and the Limits of Uninformed Learning', *Behavioral and Brain Sciences*.

Dennett, D.C. (1991), 'Mother Nature Versus the Walking Encyclopedia: A Western Drama', in W. Ramsey, S. Stich and D. Rumelhart (eds.), *Philosophy and Connectionist theory*, Hillsdale: Erlbaum pp 21–30.

Derthick, M. (1990), 'Mundane Reasoning by Settling on a Plausible Model'. *Artificial Intelligence*, 46 pp., 107–157.

Dreyfus, H.L. (1965), *Alchemy and Artificial Intelligence*, The RAND corporation paper P–3244.

Dreyfus, H.L., and Dreyfus, S.E. (1986), *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, New York: The Free Press.

Dreyfus, H.L., and Dreyfus, S.E. (1987), How to Stop Worrying about the Frame Problem even though it is Computationally Insoluable, in Z. W. Pylyshyn (Ed.), *The Robot's Dilemma: the Frame Problem in Artificial Intelligence* Norwood: Ablex (pp. 95–112),

Elman, J.L. (1991), 'Distributed Representations, Simple Recurrent Networks, and Grammatical Structure, *Machine Learning*', 7, pp. 195–225.

Elman, J.L. (1993), 'Learning and Development in Neural Networks: the Importance of Starting Small', *Cognition*, 48, pp.71–99.

Feldman, J.A. (1989), 'Neural Representation of Conceptual Knowledge', in L. Nadel, L.A. Cooper, P. Culicover, and R. M. Harnish (eds.), *Neural Connections, Mental Computation* Cambridge: MIT Press (pp. 68–103).

Fetzer, J.H. (1991), 'The Frame Problem: Artificial Intelligence Meets David Hume', in K.M. Ford and P.J. Hayes (eds.), *Reasoning Agents in a Dynamical World: the Frame Problem* London: JAI Press (pp. 55–69).

Fodor, J.A. (1975), *The language of thought*, Cambridge: Harvard University Press.

Fodor, J.A. (1987), *Psychosemantics: the Problem of Meaning in the Philosophy of Mind*, Cambridge : MIT Press.

Fodor, J.A., and McLaughlin, B. (1990), '*Connectionism and the Problem of Systematicity: Why Smolensky's Solution does not work*', Cognition, 35, 183–204.

Fodor, J.A., and Pylyshyn, Z. W. (1988), 'Connectionism and cognitive architecture', *Cognition*, 28, 3–71.

Hadley, R.F. (1994a), 'Systematicity in Connectionist Language Learning', *Mind and Language*, 9(3), pp.247–272.

Hadley, R.F. (1994b), 'Systematicity revisited: reply to Christiansen, Chater, Niklasson and van Gelder', *Mind and Language*, 9(4), pp.431–444.
Hadley, R.F. (1995), 'The 'Explicit–implicit' Distinction', *Minds and Machines*, 5, pp.219–242.
Haselager, W.F.G. (1997), *Cognitive Science and Folk Psychology: The Right Frame of Mind*, London: Sage Publications.
Haugeland, J. (1987), 'An overview of the Frame Problem', in Z. W. Pylyshyn (Ed.), The Robot's Dilemma: the frame problem in artificial intelligence pp. 77–94, Norwood: Ablex.
Haugeland, J. (1991), 'Representational genera', in W. Ramsey, S.P. Stich, and D.E. Rumelhart (eds.), *Philosophy and Connectionist Theory* pp. 61–89, Hillsdale: Lawrence Erlbaum Associates.
Hayes, P.J. (1991), Commentary on 'The Frame Problem: Artificial Intelligence Meets David Hume', in K.M. Ford and P.J. Hayes (eds.), *Reasoning Agents in a Dynamic World: the Frame Problem* pp. 71–76, London: JAI Press.
Hinton, G.E. (1990), 'Preface to The Special Issue on Connectionist Symbol Processing', *Artificial Intelligence*, 46, 1–4.
Holyoak, K.J. (1991), Symbolic Connectionism: Toward Third–generation Theories of Expertise', in K.A. Ericsson and J. Smith (eds.), *Toward a General Theory of Expertise: Prospects and Limits* Cambridge: Cambridge University Press pp. 30–335.
Horgan, T., and Tienson, J. (1996), *Connectionism and the Philosophy of Psychology*, Cambridge: MIT–Press.
Hummel, J.E., and Holyoak, K.J. (1993), Distributing Structures Over Time, *Behavioral and Brain Sciences* 16(3), p.464.
Janlert, L.E. (1987), 'Modelling change – the frame problem', in Z.W. Pylyshyn (ed.), *The Robot's Dilemma: the Frame Problem in Artificial Intelligence*, Norwood: Ablex, pp. 1–40.
Jorna, R.J., and Haselager, W.F.G. (1994), 'Associationism: Not the Cliff Over Which to Push Connectionism', *The Journal of lntelligent Systems*, 4(3–4), pp.297–309.
Kirsh, D. (1990), 'When is Information Explicitly Represented', in P. Hanson (ed.), *Information, Language, and Cognition*, Vancouver: University of British Columbia Press, pp. 340–365.
McCarthy, J., and Hayes, P.J. (1969), 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', in B. Meltzer and D. Michie (eds.), *Machine Intelligence*, Edinburgh: Edinburgh University Press pp.463–502.
McNamara, P. (1993), 'lntroduction', *Philosophical Studies*, 71, pp.113–118.
Meyering, T.C. (1993), 'Neuraal Vernuft en Gedachteloze Kennis: Het Moderne Pleidooi Voor een Niet–propositioneel Kennismodel', *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 85, pp.24–48.
Miikkulainen, R. and Dyer, M.G. (1991), 'Natural Language Processing with Modular PDP networks and Distributed Lexicon'. *Cognitive Science*, 15, pp. 343–399.
Niklasson, L.F., and Van Gelder, T. (1994), 'On being systematically connectionist', *Mind and Language*, 9(3), pp.288–302.
Palmer, S.E. (1978), 'Fundamental Aspects of Cognitive Representation', in E. Rosch and B.B. Lloyd (eds.), *Cognition and Categorisation*, Hillsdale: Erlbaum, pp.259–303.
Pollack, J.B. (1990), 'Recursive Distributed Representations', *Artificial Intelligence*, 46, 77–105.
Pylyshyn, Z.W. (ed.), (1987), *The Robot's Dilemma*, Norwood: ABLEX Publishing Corporation.
Shanahan, M. (1997), *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, Cambridge: MIT–Press.
Shastri, L. and Ajjanagadde, V. (1993) 'From Simple Associations to Systematic Reasoning: A Connectionist Representation of Rules, Variables and Dynamic Bindings Using Temporal Synchrony'. *Behavioral and Brain Sciences*,16(3), 417–494.
Smolensky, P. (1991), 'The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn', in T. Horgan and J. Tienson (eds.), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academic Publishers, pp. 281–308.
Sun, R. (1994), 'Connectionist Models of Commonsense Reasoning'. In D.S. Levine and M.1. Aparicio (eds.), *Neural Networks for Knowledge Representation and Inference*, Hillsdale: Lawrence Erlbaum Associates, pp. 241–268.
Thagard, P. (1992), *Conceptual Revolutions*, Princeton: Princeton University Press.

Van Gelder, T. (I990), 'Compositionality: A Connectionist Variation on a Classical Theme', *Cognitive Science,* 14, pp.355–384.

Van Gelder, T. (1991a), 'Classical Questions, Radical Answers: Connectionism and the Structure of Mental Representation', in T. Horgan and J. Tienson (eds.), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academic Publishers, pp. 355–381.

Van Gelder, T. (1991b), 'What is the "D" in "PDP"? A Survey of the Concept of Distribution', in *W. Ramsey*, S.P. Stich, and D. E. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Hillsdale: Lawrence Erlbaum Associates, pp. 33–59.

Van Gelder, T. (1992), Defining 'distributed representation', *Connection Science*, 4(3and4), 175–191.

Wilks, Y. (1990), 'Some comments on Smolensky', in D. Patridge and Y. Wilks (eds.), *The foundations of Al*, Cambridge: Cambridge University Press, pp. 327–336.