W.F.G. HASELAGER

On the potential of non-classical constituency

Horgan and Tienson doubt that classical cognitive science will be able to solve the frame problem but have some expectations with respect to connectionism. Since the frame problem arises in contexts where a potentially large amount of complex knowledge is involved, connectionism has to prove that its models can represent and use well-structured information. Indeed, Horgan and Tienson note that the key question for the viability of connectionism is whether representations with some kind of non-classical encoding of constituency are susceptible to richly structure-sensitive processing. I will examine several connectionist models and point out three reasons for doubting that work of this kind can scale up to the extent required for dealing with the frame problem. I will conclude that connectionism as of yet has no principled and satisfactory way of effectively representing structured information in a distributed way. Hence, the frame problem provides a difficulty to connectionism that is no less serious than the obstacle it constitutes for classical cognitive science.

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 & 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[1], 1997; Hayes, 1991; Pylyshyn, 1987), the fundamental difficulty, in my view, is that *everything* we know is potentially relevant for

[1] This paper uses some material taken from Haselager (1997) and Haselager & Van Rappard (in press). Permission by the publishers is gratefully acknowledged.

Acta Analytica 22, 1998; S. 23-42

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 I will investigate the potential of connectionism to solve the frame problem. More specifically, I 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 connectionism's potential to represent and use the structure of information has been extensively discussed in relation to the issue of systematicity, I will examine some proposed connectionist solutions to the problem of systematicity that are generally regarded as promising and that have also been advocated by Horgan and Tienson. I 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, I 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.

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 the development of non-monotonic logic (see Haselager, 1997 for a more extensive treatment of the history of the frame problem). Since the article by McCarthy & 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]. In the context of this paper, I 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.

According to a specific approach to the frame problem (that I treat more fully, together with its main alternative, in Haselager, 1997), the frame problem can be thought of as being a consequence of using a classical symbolic representational format. As is well known, Fodor (1975; Fodor & Pylyshyn, 1988) has argued that in order to explain certain characteristics of cognition that are

[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.

referred to by the terms 'productivity[3]' and 'systematicity[4]' it is necessary for the representational system to be compositional[5]. This means that representations have a combinatorial syntax and semantics, which, according to Fodor, is made possible by their *concatenative constituent* structure. This structure results 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. The concatenative constituent structure of complex representations is of direct use in representing the structure of the information represented.

On the classical view, information can have causal consequences only if it is explicitly represented by means of symbolic structures and processed by means of explicit structure-sensitive rules. The concatenative constituent structure of representations outlined above conforms to this requirement. Unfortunately, the consequence of this suggestion is that *all* the objects, constraints and relations found to obtain in the world have to be explicitly represented (e.g. by means of links between symbolic structures in a hierarchy or through the use of rules). This is problematic for if everything is explicitly represented there will be great problems in quickly locating a particular represented item. Moreover, 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

[3] *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, p. 147-148; Fodor & Pylyshyn, 1988, p. 33-37).

[4] 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.

[5] 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, p. 2-3; Pollack, 1990, p. 78; Smolensky, 1991, p. 288; Van Gelder, 1990, p. 355-356).

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.

Non-classical constituency

Horgan and Tienson (1996) adhere to a language of thought, but point 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 (p. 74). Syntax does not entail a part/whole relationship (p. 71), but merely the systematic and productive encoding of semantic relationships (p. 73). They state, correctly in my 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).

In other words the constituent need not be physically present as long as the information it carries is present ('effectively manifested') in the encompassing representation (p. 80). Moreover, the representations are not processed by structure sensitive rules but exert their influence as defeasible causal tendencies (based on their location in a well-molded, inclination-rich landscape). As far as I can see then (but see Markič (this volume) for a more extensive treatment of Horgan & Tienson's views on the Language of Thought), they agree with proponents of the Language of Thought hypothesis about *what* a representational format has to be able to achieve (i.e. encode causally effective constituent structure in representations), but they thoroughly disagree about *how* this has to be achieved. Syntax is necessary but it need not be based on concatenative constituency.

In a similar vein, Van Gelder (1990, 1991a) has suggested that compositionality can be achieved *without* a concatenative constituent structure. According to Van Gelder, a representation can represent a structured item without itself having a constituent structure, as long as there are general, effective, and reliable processes by which to compose complex expressions from constituents and to decompose a complex expression back into its constituents. Processes are general if they are applicable to arbitrarily complex representations, effective when they can be performed mechanically, and reliable if they always generate the same answer for the same input (Van Gelder, 1990, p. 361). He claims that distributed representations[6] can comprise a functionally compositional representational scheme that is at least in principle capable of representing and using the structure inherent in information. Points in an activation or weight space are standard examples of distributed representations. However, designing and implementing proper compositional and decompositional processes is no easy matter. It is even more difficult to create a model that is capable of a richly content-appropriate processing of the representations thus created. In this respect, I find Horgan and Tienson's discussion of the 'moving target' strategy most interesting.

The moving target strategy

Horgan and Tienson (1996, p. 7) view cognitive processes as the effect of cognitive forces emitted by representations. From a connectionist perspective a representation can be understood as a point or region in an activation space[7] (p.

[6] 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 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).

[7] However, Horgan and Tienson point out that not every point in an activation landscape is considered to be a representation: "total cognitive states are mathematically realized by points in high-dimensional activation space: from the perspective of the mathematical level of description, these points are the representations. (…) Not every point in activation space need be a representation; in fact, typically, very few will be." (p. 65). Typically, points that constitute representations will be attractors: "Representations must be attractors because not all physical states of a cognitive system are cognitive states, and cognitive states must evolve from cognitive states to cognitive states." (p. 195, n. 6).

65). The relations between points in an activation landscape can embody semantic information inherent in cognitive states, their interrelations and their constituents (p. 149, 156). Cognitive forces actively dispose the cognitive system toward a certain cognitive outcome (p. 150). This can be understood from a dynamical perspective as the slope or incline on the activation landscape leading to another point in the landscape (p. 151). That is, inclines in an activation landscape constitute cognitive forces (p. 167). Importantly, the direction of cognitive forces is determined by the structured content of the representations:

> "Each token cognitive state undergoes its type of force-determining process because of its semantic constituents and their relations." (p. 102, see also p. 98).

Horgan and Tienson suggest that automatic and systematic content-appropriate cognitive transitions are the joint product of a subtly contoured activation landscape and a subtle realization relation from total cognitive states to points on that landscape (p. 67). Cognitive forces and non-classical representations can jointly produce this result by means of a training regime called the *moving target* strategy. Thanks to the moving target strategy, both the activation landscape as well as the realization of intentional states as points on this activation landscape are changed. The weights, determining the shape of the activation landscape, and the activation-vectors that 'count' as representations are set accurately for the task at hand by the moving target strategy. By allowing the activation vectors realizing representations to change one can strategically locate the representations on the activation landscape instead of allocating them arbitrarily. Thus, a controlled co-evolution of the activation landscape and the realization for all representations can be achieved (p. 154).

> "The realization relation exhibits increasing systematicity, coming to reflect, in the way it positions representation-realizing points relative to one another on the activation landscape, important relations among the intentional states being realized. The realization relation and the landscape topography end up "made for each other" with respect to the information-processing task the system is being trained to perform: the final weight setting for the network subserves a high-dimensional activation landscape whose overall local topography yields systematically content-appropriate temporal trajectories under the operative intentional/mathematical real-

ization relation. Thus, the key to the system's design is that the shape of the activation landscape and the overall positioning of the representation-realizing points on that landscape are jointly just right to subserve the relevant intentional transition function for a very large class of potential intentional states." (p. 62).

It should be noted that Horgan and Tienson explicitly mention the frame problem as one of the reasons for dismissing classical cognitive science. Their acceptance of noncomputational dynamic cognition implies that they expect the frame problem to be solvable this way (e.g. see p. 67 where they promise to explore the potential of the nonclassical framework to overcome the problems for classicism).

I think the basic idea of Horgan & Tienson in relation to the frame problem is clear: Automatic content-appropriate interaction between cognitive forces can be subserved by superimposed cognitive inclines between well-located representational points in a high-dimensional activation landscape (p. 153). One may think of the weights of a network as reacting to input by creating a landscape that will automatically 'provide' representational points with their 'proper inclinations'.

The non-concatenative constituently structured representations allow for a content-appropriate but very flexible and defeasible interaction (i.e. cognitive-state transition functions that need not be tractably computable). Thus, the frame problem might be solved or circumvented.

It bears emphasis that this proposed solution to the frame problem is made possible precisely because connectionism forswears the symbolic representational (i.e. classical concatenative constituency) format. Knowledge need not be represented explicitly nor processed by structure-sensitive rules. Instead, the knowledge of a network, embodied in its weights, 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 ('landscape creation') the network is capable of.

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 and models of connectionism 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 & Pylyshyn, 1988; Holyoak, 1991, p. 315-316; Thagard 1992, p. 242-243).

Horgan and Tienson claim that there are examples of a rudimentary kind of structure-sensitive processing of non-classical constituently structured representations. They are, furthermore, quite optimistic with respect to the potential of connectionist models to preserve structure in the representations:

> “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 Pollack, Berg, Chalmers and Smolensky as examples of the kind of nonclassical syntax they endorse. However, in my opinion it is far from clear whether the models proposed actually succeed in capturing structure and displaying systematicity. Furthermore, I will indicate reasons to doubt that the mechanisms used allow a scaling up to realistically complex contexts. The question, then, is whether syntactic sensitivity is possible to a realistically complex extent without concatenative constituently structured (part/whole) representations?

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 non-concatenative representations.” (Van Gelder, 1990, p. 381; see also Chalmers, 1993, p. 312; and Fodor & McLaughlin, 1990, p. 202, n. 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[8]. The resulting distributed representations are used by the actual transformation network (a basic, three layer feed-forward 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

---

[8] 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, p. 55-56).

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.

Lawfulness versus coincidence

Importantly, Horgan and Tienson (1996, p. 80) claim that the existence and useability of non-classical 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[9]. 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, I 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

[9] 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.

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 & McLaughlin, 1990, p. 202-203; Fodor & 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[10].

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. I 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, I suggest, most beneficially be seen as an attempt to provide a safeguard against too readily taking ‘accidental’ signs of systematicity for the real thing. I propose, then, to take the requirement of displaying systematicity under many different

[10] 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 & 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, p. 252-254, p. 270).

(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 & 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 my 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 my 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 & 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[11]). It seems unavoidable that the amount of training needed will become unmanageable in

[11] Christiansen & 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

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[12]). 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 qˇp 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, I 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'[13]. As they say:

> "The *design* and *training regime* of the representation generator results in representations that are systematically positioned in the

[12] The point I am 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 (1997) 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. My point is that most details of the training regimes lack such justification.

[13] 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, p. 296-297).

> space so that the representation for 's' occupies the space in between the 'known' constituents" (Niklasson and Van Gelder, 1994, p. 297-298; my emphasis; see also Hadley, 1994b, p. 437-438).

As in the case of Christiansen and Chater, I 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, I 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, I 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 I fail to see how connectionism would be able to deal with that problem successfully. Before drawing my final conclusions, I 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.

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 common-sense 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), Shastri & Ajjanagadde (1993), and Sun (1994).

For instance, a recent and ingenious connectionist model of common-sense reasoning is outlined by Shastri & 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. I 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.

Conclusion

The frame problem is generally regarded as a serious, sometimes even unsolvable, difficulty for classical cognitive science. The fact that the symbolic representational format allows for a distinction between what is explicitly and implicitly represented can be thought of as the underlying cause of the problem. From such a perspective, distributed representations can be seen as promising.

However, although I consider the potential of distributed representations to be most interesting, I 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. I 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, I 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. In all, Horgan and Tienson's expectations may not be justified.

REFERENCES

Butler, K. (1993): Connectionism, classical cognitivism and the relation between cognitive and implementational levels of analysis, *Philosophical Psychology, 6* (3), 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), 305-319.

Christiansen, M.H., & Chater, N. (1994): Generalization and connectionist language learning, *Mind & Language, 9* (3), 273-287.

Churchland, P.M. (1989): *A neurocomputational perspective: the nature of mind and the structure of science*, Cambridge: MIT Press.

Clark, A., & Thornton, C. (1997): 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 & D. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 21-30), Hillsdale: Erlbaum.

Derthick, M. (1990): Mundane reasoning by settling on a plausible model. *Artificial Intelligence, 46*, 107-157.

Dreyfus, H.L., & 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* (pp. 95-112), Norwood: Ablex.

Elman, J.L. (1993): Learning and development in neural networks: the importance of starting small, *Cognition, 48*, pp. 71-99.

Fetzer, J.H. (1991): The frame problem: artificial intelligence meets David Hume, in K.M. Ford & P.J. Hayes (Eds.), *Reasoning agents in a dynamical world: the frame problem* (pp. 55-69), London: JAI Press.

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

Fodor, J.A., & McLaughlin, B. (1990): Connectionism and the problem of systematicity: Why Smolensky's solution does not work, *Cognition, 35*, 183-204.

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

Hadley, R.F. (1994a): Systematicity in connectionist language learning, *Mind & Language, 9* (3), 247-272.

Hadley, R.F. (1994b): Systematicity revisited: reply to Christiansen and Chater and Niklasson and van Gelder, *Mind & Language, 9* (4), 431-444.

Haselager, W.F.G. (1997): *Cognitive science and folk psychology: the right frame of mind,* London: Sage Publications.

Haselager, W.F.G., & Van Rappard, J.F.H. (in print). *Minds and Machines.*

Hayes, P.J. (1991): Commentary on "The frame problem: artificial intelligence meets David Hume", in K. M. Ford & 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 & J. Smith (Eds.), *Toward a general theory of expertise: prospects and limits* (pp. 301-335), Cambridge: Cambridge University Press.

Horgan, T., & Tienson, J. (1996): *Connectionism and the philosophy of psychology,* Cambridge: MIT-Press.

Hummel, J.E., & Holyoak, K.J. (1993): Distributing structures over time, *Behavioral and Brain Sciences 16* (3), 464.

Markič, O. (1998): Connectionism and the language of thought, *Acta Analytica*.

McCarthy, J., & Hayes, P.J. (1969): Some philosophical problems from the standpoint of artificial intelligence, in B. Meltzer & D. Michie (Eds.), *Machine Intelligence* (pp. 463-502), Edinburgh: Edinburgh University Press.

McNamara, P. (1993): Introduction, *Philosophical Studies, 71*, 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*, 24-48.

Niklasson, L.F., & Van Gelder, T. (1994): On being systematically connectionist, *Mind & Language, 9* (3), 288-302.

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.

Shastri, L. & 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 & J. Tienson (Eds.), *Connectionism and the philosophy of mind* (pp. 281-308), Dordrecht: Kluwer Academic Publishers.

Sun, R. (1994): Connectionist models of commonsense reasoning. In D.S. Levine & M.I. Aparicio (Eds.), *Neural networks for knowledge representation and inference* (pp. 241-268). Hillsdale: Lawrence Erlbaum Associates.

Thagard, P. (1992): *Conceptual revolutions*, Princeton: Princeton University Press.

Van Gelder, T. (1990): Compositionality: a connectionist variation on a classical theme, *Cognitive Science, 14*, 355-384.

Van Gelder, T. (1991a): Classical questions, radical answers: connectionism and the structure of mental representation, in T. Horgan & J. Tienson (Eds.), *Connectionism and the philosophy of mind* (pp. 355-381), Dordrecht: Kluwer Academic Publishers.

Van Gelder, T. (1991b): What is the "D" in "PDP"? A survey of the concept of distribution, in W. Ramsey, S. P. Stich, & D. E. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 33-59), Hillsdale: Lawrence Erlbaum Associates.

Van Gelder, T. (1992): Defining 'distributed representation', *Connection Science, 4* (3&4), 175-191.

Wilks, Y. (1990): Some comments on Smolensky, in D. Patridge & Y. Wilks (Eds.), *The foundations of AI* (pp. 327-336), Cambridge: Cambridge University Press.