A PRAGMATIC LOOK AT ARTIFICIAL INTELLIGENCE or: The proper proper treatment of connectionism *

Jacob L. Mey

Odense University

Based on recent claims and counter-claims about the nature of human mental processing, this article reviews some of the evidence presented, and tries to dispel some myths surrounding the 'new' cognitive science (also called 'PDP', 'connectionism', 'neural network theory', and so on).
In particular, the question of the so-called 'subsymbolic' level of representation is raised, and some of the implications of implementing fully connectionist machines in a human surrounding are discussed.

0. Introduction: A new deal in AI

Recently, a new development has struck the field of AI (and in general, of computational data processing, 'Informatik', as the Germans call it.) The development is called connectionism (sometimes, and more or less equivalently in certain dialects of AI-ese, also known as PDP (for 'parallel distributed processing') or referred to as 'neural network' theory -- the distinctions between the different terms are not trivial, but need not occupy us here).

What is important about connectionism, and why does it have such an impact on AI? Recently, I came across a strong formulation of the link between AI and connectionism: "Konnektionismus ist Künstliche Intelligenz", i.e. 'Connectionism is AI' (Diederich 1988:28, as quoted by Hoeppner 1988:27). Implying that, as Hoeppner seems to do, the two are simply identical (which indeed is one possible reading of the copula is), seems an unreasonable interpretation of what, in the context, could have been no more than a programmatic statement, or maybe even better, a wishful thought.[1] But that there is a strong link between connectionism (as a way of looking at, and practicing, electronic data processing, to take a rather weak interpretation of the term)
and artificial intelligence, seen as an endeavor to unite different areas of human research in a

computational environment and philosophy, is beyond doubt (as also Hoeppner has noted (ibid.)).

What I would like to do in the following is: first, characterize connectionism as a 'program of research' (Lakatos' term; cf. Winograd & Flores 1986:24) within (or even encompassing) AI; then, talk about some of the difficulties that arise when one tries to incorporate the traditionally accepted notions of the human sciences (such as psychology, linguistics, human learning theory, etc.) into what some have called a new 'paradigm' of understanding (in the sense of Kuhn 1962), but what to others is no more than a novel way of implementing on the computer (albeit not always successfully) earlier acquired knowledge and insights.[2)]

1. The Black Box and Other Myths

Ever since the days of Chomsky (1957), the idea of a 'Black Box', i.e. a device that was able to perform certain operations without caring about their content (and without knowing either, for that matter) has been popular among linguists. And even though Chomsky himself has always (and rather vigorously) protested against the notion that his grammatical model was influenced by the computer, or even directly inspired by computational methods, it seems clear, in retrospect, that that was exactly what was happening in the early days of TG. And furthermore, that the computer connection has been responsible for much of the early success and popular appeal of TG.

A variant of the 'Black Box' is the 'Chinese Room', made famous in discussions on AI inspired by Searle's seminal lectures from 1984 ('Minds, brains and science'; originally a series of BBC radio talks, 'The Reith Lectures'). This particular version of the 'black box' is distinguished by the fact that its (Chinese) input and output are manipulated by a robot inside the box ('the Chinese Room'); that the robot doesn't know anything about the semantics of the language ('Chinese') it is manipulating; and that it strictly executes certain reshuffling operations (equivalent to syntactic rules) only on the symbols of the language, or their equivalents within the room (their representations).

What is important here is that within the box, we can talk about symbols or representations indiscriminately: it doesn't make the slightest difference to the robot whether the things it does within the 'Chinese Room' have to do with Chinese characters, or some constellations of 'room-units' (e.g. blinking lights, blocks with names on them, and so on), as long as the output is right in relation to the input. In other words, the robot doesn't have to understand Chinese in order to be able to manipulate the rules for Chinese 'correctly', i.e. according to the syntax of Chinese. As to the semantics of Chinese, or the content of those symbols, it couldn't care less anyway.

However, in this kind of reasoning, once you've said A, you have to say B as well. For, why stop at symbols? Underneath the symbolic level, we have that of the 'subsymbolic', for instance what Hjelmslev (1943) has called the level of 'glossemes' (units of linguistic analysis below the sign boundary), or what in phonology has become known as the set of 'distinctive features', units of analysis that (within certain limits) could be combined freely across traditional, symbolic boundaries, such as those between vowels and consonants, and so on. If the 'Black Box' is all about operations and manipulations, and if the Chinese robot doesn't care what the units of manipulation look like, or 'are about', then why should we care, as long as the outcome is right? All's well that ends well, and the proof of the pudding is in the eating -- not in its making.

There is, of course, a traditional rub. How are we going to deal with those subsymbolic units? In the case of phonology, despite the 'universal', abstract character of the distinctive features, we still have some 'feel' for the quality denotated ('represented') by features such as [strident] or [grave]. (For example, in the case of the former, we would have some trouble assigning it to a vocalic, rather than to a consonantal member of the phonetic inventory.) So, in a way, we play it both sides: we pretend to be an ignorant robot, obeying only formal rules, but at the same time we know damn well why we are doing what we're doing, and can also remember what we're doing, so we can repeat it time after time. And suppose we pretend to stand wholly aside, leaving everything to the robot or the machine on the subsymbolic
level, there is still the problem of sequencing: Which of the features (or other subsymbolic items) goes with what other(s), and in what order?

So, what we're dealing with is not only the Black Box itself, but a whole series of connected notions, of which that of the subsymbolic is an important, but by no means unique exemplar. Like in the case of the 'Black Box', one has to distinguish between what is real and what is not, with regard to those notions and their associated claims. In the following, I will deal mostly with what I will call the 'myth' part: myths about AI that are revived, highlighted, and reinforced in and through recent 'connectionist' forms of thinking. For each of these, I will argue as follows: Admitting that there is some truth and usefulness in some of these notions, it still is important to separate out the real part from the ideal or idealized one (and, a fortiori, from the 'myth'). As to the claims made, in many cases, these belong more to the mythical than to the real; and besides, they may not even be necessary, from a pragmatic point of view. The notions (or myths) I will discuss below are: the subsymbolic vs. the symbolic; the brain and its 'architecture'; parallel (distributed) processing; and finally, implementational performance.

2. The subsymbolic

In some exposés of connectionist theory (e.g. Smolensky 1988), the point is made that the main difference between what could be called 'old AI' and 'new AI' (Hoeppner 1988:28) is the introduction of the 'subsymbolic' level. Whereas in the classical account (see, e.g., Fodor & Pylyshyn 1988), computation is thought of as a manipulation of symbols on a (digital) machine, the '(counter-)revolutionary' view ascribed to the Connectionists[3] is that we do not need any independent level of symbolic representation, or indeed any representation at all: everything can be dealt with at the level of the subsymbolic (if indeed we need to speak of 'level' here). Rather than dealing with traditional macro-units (such as, e.g. in phonology, the phonemes of a language), a connectionist account defines low-level, subsymbolic features that cannot strictly be said to represent any symbol at all, at least not directly. Building up those (after all, indispensable) higher units (symbols) happens through associative connections between the subsymbolic 'nodes', in accordance with the weight of the input, distributed over the entire network. (I will not go into details here; for a clear description of how these subsymbolic features ('Wickel-features') work, see Pinker & Prince (1988)).

So, in a way one could say that networks with distributed, weighted associative connections have come to replace the old, hierarchically organized, constituency-defined structures. Under this interpretation, the difference between the two approaches would be one of architecture, of building blocks of different kinds, some larger, some smaller, and of the way to put them together: on top of each other, or in a randomly organized jumble that is retrievable only by means of a computer program. The things that both the symbols and the subsymbolic units represent, would be the same: "both the conceptual and the sub-symbolic levels postulate representational states, but sub-symbolic theories slice them thinner" (Fodor & Pylyshyn 1988:9).

However, as Dietrich & Fields have pointed out (1988:30), the debate is not so much about architecture as about the interpretation of the structures generated by the different architectural models. In other words, we are dealing with a semantic controversy: What do the symbols, respectively the subsymbols, represent? For mentalists, such as Fodor or Pylyshyn, there is no doubt: the symbols belong to a language of the mind, a 'mentalese' (Fodor 1975), whereas subsymbolic entities don't, which is why they're useless in describing and explaining human activity. For Smolensky, on the other hand, symbols are useless (or too complicated to deal with; see his diatribe against symbolism (1988:4-6)); subsymbolic units only are able to provide the precision, formality, and completeness that is needed to simulate human 'intuitive processing' of mental data.

But not only do we not need a 'conceptual' (read: symbolic) level of description, the 'subsymbolic hypothesis' that is said to be "the cornerstone of the subsymbolic [read:

connectionist] paradigm" (ibid.) is stated negatively as:

"The intuitive [human] processor is a subconceptual connectionist dynamical system that does not admit a complete, formal, and precise conceptual-level description." (Smolensky 1988:7).

However, if we take Dietrich & Fields' remark (1988:30), quoted above, about the semantics seriously, and at the same time recall that Smolensky operates within a physicist paradigm, it may be tempting to try and reconcile the two 'levels' (as also suggested by Dietrich & Fields, in continuation of Smolensky's own proposal (1988:12)) by assuming that there is a smooth transition between the subsymbolic and the symbolic, and that connectionism is nothing but a "microtheory of cognition that stands to macroscopic cognitive science as quantum mechanics stands to classical mechanics" (Dietrich & Fields 1988:30). This would, in fact, amount to interpreting lower-level units as in some sense assignable to higher-level ones. The representation relation may not be a one-to-one mapping or a strict, hierarchical constituency (as claimed by Fodor & Pylyshyn); however (to continue the physics example), everybody will agree that quarks, e.g., belong to the sub-atomic level, yet, there is some sense in which they are assigned to the atom: the two levels are not incompatible, and their respective representations are determined by the 'semantics' of each level. As Dietrich & Fields remark: "The only restriction on the semantics of the interpretations used to describe the system are those imposed by intra-level coherence, and by the explanatory goals with which the interpretation is constructed. One can, if one wants, interpret neurons as representing grandmothers; if this interpretation does not prove to be useful, it can always be revised." (Ibid.:31)

Still, on Smolensky's and other connectionists' views, such as 'ecumenical' interpretation cannot be tolerated. The reason is that the symbolic and subsymbolic paradigms are basically and irrevocably incompatible: one cannot be reduced to the other, and they are mutually inconsistent, not in the sense that there couldn't be a way of implementing one program in terms of the other, syntactically, but because of 'real' differences.

According to Smolensky, there is something which the subsymbolic level provides that cannot be captured at the symbolic level, viz. "a complete formal account of cognition" (1988:7). However, the vexing problem still is with us how to map that formal account onto everyday concepts of cognition, i.e. the concepts humans use to deal with their world. While mentalists claim that there is a conceptual (symbolic) level that describes
human cognitive activities satisfactorily, or even necessarily, connectionists claim that this is not the case, at least not in the sense of a formal, precise, and complete description. But (pace Fodor) how are we to know that the precise, formal, and complete account that we have obtained on the subconceptual level, thanks to connectionist methods, indeed represents human cognitive activities at the symbolic level? Or (to turn the argument in the other

direction, as suggested by Dietrich & Fields (1988:30)), how are we to know that the subconceptual level postulated by the connectionists indeed is a formal, etc. representation of what is going on at lower levels, such as the neural, the biochemical, etc.? And I quote: "If there is something to the claim that concept-level descriptions are fuzzy in principle, what prevents us from using the same argument to show that subsymbolic descriptions are only fuzzy approximations of biochemical descriptions, and so forth?" (Ibid.:30) Isn't being a conceptualist or a mentalist just as bad as being a 'sub- conceptualist' or 'sub-mentalist' -- if indeed, there are no other weighty arguments around than the ones suggested by the respective protagonists of those schools?

In the following, I will examine precisely such an argument. It revolves around the old question of the workings of the brain, and how to explain them.

3. Connectionism and 'brainware'

> "If the human brain were so simple that we could
> understand its workings, then we would be so dumb
> that we couldn't".
>
> (Graffito (Rees 1983))

Being no expert on neurophysiology or on 'brainware', in what follows, I will try to avoid the pitfalls alluded to in the above quote. That is to say, I accept prima facie the evidence about the relative slowness of the human brain, as compared with digital computers, and the ensuing need for some kind of parallel computing in the brain (cf. the 'hundred step' constraint as defined by Feldman et al. (1981, 1982)).

What is at stake here, however, is not the factual implementation of human neurological activity in the brain, but the conclusions that some have drawn from this activity, and the
arguments that are built around those conclusions in order to prop up certain connectionist claims.

Connectionism, it is often said, is descriptively more correct than the 'classical' theory (whatever is meant by that), because it "as it were, sneaks up on the brain itself and cribs its tricks" (thus, more or less, Hoeppner 1988:28). Others (e.g. Fodor & Pylyshyn 1988:62) talk about 'brain-style modeling', which, in a very wide sense, may be taken to mean that "theories of cognitive processing should be influenced by the facts of biology" (a tenet with which most of us would agree, presumably); alternatively, and in a more (or even very) narrow sense, 'brain style modeling' can be taken to comprise the explicit modeling of human cognitive activities on "properties of neurons and neural organizations" (cf. 1988:62).

Whatever interpretation of this 'modeling style' is chosen, for a mentalist it seems a

priori and intuitively clear that cognitive activity is more than circuits opening and closing, be they neural in character or electronic. The connectionist, of course, is not willing to admit such intentions as evidence, but surely the mentalist is entitled to evidence for the connectionists' claim that indeed the connectionist models are better because they emulate the structure of the human 'brainware', as postulated, and in part corroborated, by neurophysiological research? After all, the debate is not about hardware (inclusive 'brainware'): what we're interested in, is the way that hardware (or 'brainware') is put to work, is 'programmed', to use a computer metaphor.

The point here is not whether or not the brain is structured 'like' a computer (serial or parallel). The facts of the 'brainware' and its organizational patterns can provide interesting and important information about the way we go about our business using that brain. Yet, brain activity is not exhaustively limited to, or described as, the way the neurons 'fire' or the synapses are activated. In any case, a postulate to that effect (such as subscribed to by many connectionists) is no more than that, until concrete evidence has been put forth excluding all other possible interpretations. In particular, it seems doubtful whether the neuronal organizational level should be incorporated in toto, and as such, into the organization of our thoughts and our concept-forming activities. There is no a priori motivation for assuming any one-to-one correspondence between the two, and (as Fodor and Pylyshyn remark), "the structure of 'higher levels' of a system are [sic] rarely isomorphic, or even similar, to the structure of 'lower levels' of a system" (1988:63).

Notice that I'm not saying that it is impossible to imagine a 'brainware'-oriented model of the mind; neither can it be denied that for some areas of neuronal activities, there are structured correspondences or even analogies (and who knows, perhaps isomorphisms) between the two levels of organization (vision and motor control may be such areas; cf. Fodor & Pylyshyn, ibid.) What I am saying is that there is neither an a priori guarantee that this is so, nor a logical or psychological necessity that it must be so; and that, therefore, the brain-mind analogy cannot be used as a supportive argument for the theory that precisely presupposes such an analogical, or even isomorphic, structure. To use a somewhat trite analogy: the fact that the knee-jerk reflex is sufficiently explained by referring to sensori-motoric connections in the spinal cortex does not necessarily entail that all human nerve-activity should preferably or uniquely be explained without reference to the 'higher' processing level of the brain itself. The question whether or not "neural networks offer a 'reasonable basis for modeling cognitive processes'" (Rumelhart & McClelland 1986:110; cf. Fodor & Pylyshyn 1988:68) is clearly an empirical one and cannot be decided a priori either by mentalists or by connectionists.

4. PDP and neural networks

In discussion on connectionism, much has been made of the distinction between 'old' style computers (also called 'sequential' or 'serial' machines) and 'new'-style ones (parallel or 'Non-Von' machines, as different from the classical 'Von' (Neumann) ones). What seems beyond doubt is that a model of the brain that wants to come close to the latter's actual 'architecture' will have to be based on parallel, distributed processing ('PDP': I'll leave out the technicalities).[4] Connectionist machines do nothing but implement this new architecture, in which knowledge representation no longer is a
matter of devising the right symbolic framework, and then inputting the symbols by means of classical (e.g. LISP-style) programming, but rather, by activating a network of highly interconnected units whose total patterning is said to 'represent' (in some weak, and specifically non-retrievable way) the original input.

It has been shown that such networks have been remarkable (and indeed, surprising) properties when it comes to learning, reasoning, and in general, computing at advanced levels. What is more doubtful, however, are the generalized claims about those networks (especially in their 'neural' variety), according to which they should be able to imitate and emulate human cognitive performance tout court.

Looking at the available evidence, one is struck by the fact that learning in a connectionist environment mostly has to do with the acquisition of properties for which relatively simple rules (e.g. of a syntactical nature) are available. Such is the case for the widely publicized abilities of the connectionist program due to Rumelhart & McClelland (1986b), purported to be superior to a traditional model, both insofar as the program gives a formal computational description of the correct forms of the English past tense, and as it provides an explanation of the process of acquisition of those forms by human learners (in particular, developing speakers).

On balance, it seems that the claims made by the connectionists, viz., that their model provides a superior account compared to the traditional, rule-based approach, and that developing speakers' acquisition of a particular linguistic competence, such as knowing how to inflect strong and irregular verbs in English, is more realistically modeled in their framework that in any previous, non-connectionist one, are not entirely borne out by the empirical facts. As Pinker & Prince note (1988:164), the PDP model is at best no better than, and in a number of cases inferior to, the traditional one as far as the actual account is concerned. As to the developmental aspect, one could imagine an equally explicit, rule-based account that would work at least as well. In fact, Pinker & Prince provide such an account in a sketch-like, but still rather detailed form (1988:128-165), incorporating two of the main features of the PDP
model: sensitivity to frequency of occurrence (computed on the type, not on the token of the

verb form, and combined with stronger weighting of more likely forms), along with a competition among various candidates, resulting in the right form being preferred. Pinker & Prince note that these properties by no means are unique to PDP models, and in fact are independent of them (ibid.:130). To this, compare Hoeppner's comment: "...we are dealing here with typical pattern recognition problems, viz. the recognition of syntactically describable regularities in the input data ... a syntactic or pragmatic level cannot be detected in these [PDP] systems." (1988:30).

In my opinion, connectionists and other researchers will profit by taking comments and criticism such as those quoted above seriously. The ultimate test of any interpretive process, be it symbolic or non-symbolic, must be its descriptive adequacy (not to speak of explanation). Pinker & Prince have gone to great lengths to evaluate the PDP proposal "in terms of its concrete technical properties rather than bland generalities or recycled statements of hopes or prejudices" (ibid.); the same holds for another criticism of the past tense learning program, that by Lachter and Bever (1988). In particular, these authors raise the issue of to what extent the amazing results of the connectionist learning program are due to special effects that are introduced ad hoc, such as the sharpening of boundaries between the individual Wickelphones (1988:209)[5], and they generalize this observation to something they humorously, but not entirely without malice, dub 'TRICS', viz., "'The Representations It [the PDP model] Crucially Supposes'" (ibid.:208).

Lachter and Bever conclude their discussion of three PDP models of learning (including Rumelhart & McClelland's) by stating: "... we have shown that both the learning and adult behavior models contain devices that emphasize the information which carries the rule-based representations that explain the behavior." (1988:233). In other words, if you need rules anyway (and by implication, a symbolic or conceptual level of explanation and description), then why go to all the bother to avoid them explicitly? And surely such non-rule based models cannot claim to be a sufficient and necessary replacement for older, rule-based approaches.

5. Performance and the pragmatic view

The ultimate test of any model must, of course, be its applicability to serious problems or issues of human life and existence. In this sense, the claims made by the connectionist school purport to deal with real-life matters: How do we, for instance, explain language learning? And more generally: How well can the model explain the workings of the human mind -- the overall issue in cognitive science? Or, put in other words: what kind of human mind are we envisioning, using connectionist models, and can we deal also with other, broader issues such as human responsibility, the realities of societal life, and so on? How well can the PDP model be supposed to perform from a pragmatic point of view?

In this section, I will first deal with some of the problems that have to do with successful implementation of the connectionist proposals, seen as models of the human mind. Fodor & Pylyshyn remark that connectionist models, like the older associationist ones, are not sensitive to structure, but only to frequency. Thus, the strengthening of the network connections that explain the learning process is operated in accordance with a statistical metric of co-occurrence of certain stimuli and certain responses (Fodor & Pylyshyn 1988:31). Similarly, Hoeppner remarks that connectionism, as other behaviorisms, relies on "associative atomism" (1988:29). A model of the human mind that is based on this philosophy, no matter how implemented, will have to deal with the classical objection against behaviorism, viz., its lack of cognitive adequacy (for instance, as formulated in Chomsky's critique of Skinner (Chomsky 1959)). We are left with "a gnawing sense of deja vu [sic]", as Fodor & Pylyshyn conclude their article on 'Connectionism and cognitive architecture' (1988:69).

A more general line of reasoning about cognitive adequacy would incorporate aspects that are usually referred to as pragmatic, i.e. having to do with the users of an implemented model, connectionist or otherwise. First of all, it should be clear that the main goal of cognitive research (and AI as a specific instance) should be to get to know the human mind and understand its workings, not to produce a working replica of what humans are supposed to be at their best. Why have a robot compose
a symphony that is just as good as Beethoven's Third, as long as we have (or have had) Beethovens around that are (or were) perfectly capable of taking care of such tasks?

Current fantasies about AI and its 'applied' offshoot, Expert Systems, tend to focus on the role of computerized technological systems that will be able to take over a large part of humans' traditional tasks in engineering, planning, diagnosing, repairing, and so on, in the most varied realms of human activities. In particular, the military's interest in automated weapons systems is well known, its latest manifestation being the notorious Strategic Defense Initiative, also known as 'Star Wars'. All these systems pose questions of implementability and practicality; and they force us to rethink familiar notions such as reliability, decision procedures, and so on.

In a recent book, Stuart Dreyfus tells a refreshing anecdote, illustrating the gap between dreams and realities in this domain: When asked to explain the principles along which expert systems work, he used to come up with the example of buying a car. Suppose, he said, you want to buy a new car. Wouldn't it be nice to have all the necessary information stored in a system that would tell you not only what kind of automobiles were available that corresponded to your specifications, but also all the technical details: repair costs and availability of parts, road performance, supposed or allowed depreciation, and so on? A system that, in addition to all that, suggested a scheme for financing the operation, complete

with competitive bids from several money providers, specifying the details of credit, repayment of loan, and so on? A system that you, after having considered all this evidence, could ask to make the decision, and actually buy your car?

Usually, Stuart's story went down well at parties and other social gatherings where one inevitably is confronted with the question: "And what do you do?" -- a question we all know and dread, and to which Stuart thought he had a standard, satisfactory answer. Until one day a young woman asked him: "Dr. Dreyfus, and is this the way you decide when to replace **your** car?" To which Dreyfus replied that he couldn't dream of such a thing -- buying a new car to him was much too important to be left to a mathematical model or a machine! (Dreyfus & Dreyfus 1986:9-10)

In connection with the 'new' trends in AI, the innocuous question of Dreyfus' party conversationalist takes on a wholly new aspect, too. One of the main differences between the 'old' and the 'new' model is that the latter does not respect, or recognize, the level of symbolic structures at input, and that subsequent activations throughout the network at no point resemble those structures, as they are represented in the mind. That means (and connectionists make a point of stressing this as the hallmark of their system) that the path of the activation through the network is basically not recoverable. Thus, the system may be performing correctly, but we don't know how it got at its correct results. While the program is learning, the weightings of its internal connections are changed, but since there are, in principle, no connections that we can trace to the original input, we don't exactly know what it is learning, and how and where it is modifying itself. In other words, "even if a connectionist system manifests intelligent behavior, it provides no understanding of the mind because its workings remain as inscrutable as the mind itself" (Shepard 1988:52). And this lack of understanding means, concretely, that we are unable to correct the system from the 'inside', so to speak, locating the error by inspection. For a connectionist, the system's only mode of interaction is 'take it or leave it': you can't fight statistics, the 'Black Box' reigns supreme. As Lehnert has put it, quite to the point, in my opinion: "If the connectionists ever should come to dominate AI, we will have to deal with the very real possibility that we might be able to simulate something without understanding it very well." (1987:3; cf. Hoeppner 1988:28)[6]

The localization problem alluded to here is by no means unfamiliar to connectionists such as Smolensky; cp.:

"... failures of the system to meet goal conditions cannot in general be localized to any particular state or state component. In symbolic systems, this assignment of blame (Minsky 1963) is a difficult one, and it makes programming subsymbolic systems by hand very tricky." (1988:15)

By the same token, however, if anything goes wrong, it will be difficult to deal with a

potential emergency, invoking traditional concepts of diagnosis and repair. This may not be too much of a problem as long as we are at the level of the laboratory
experiment and try to teach the computer the past tenses of the English verb or some other, undangerous knowledge, but what could happen if the program were trained (or better: had trained itself) to intercept an enemy airplane and -- upon due recognition but failing somehow to pick up on the necessary and sufficient clues
-- shoot it down? With the recent Iranian airliner tragedy in mind, this kind of danger is not at all illusory.

Here, Hoeppner's thoughtful remarks about 'Connectionism and social reality' (1988:2.5) deserve attention, when it comes to assigning the responsibility for the proper functioning of connectionist systems:

"Connectionist systems are self-organizing systems, i.e. systems that adopt to their environment in ways that are not predictable from the outside; neither can they be steered intentionally. ... in the last instance, there is nobody who could be held responsible for the system's actions (with the exception perhaps of the person who pulled the switch...)". In regular programming, Hoeppner continues, "there is a causal chain between the elements of the program and its results, at least initially and in principle; here, it is still possible to discuss matters of responsibility. When it comes to connectionism in its extreme variety, however, such an embedding in existing social and ethical contexts is hardly possible, unless one were to redefine those contexts, and by implication, ourselves." (1988:30)

Indeed, a lot has still to be said on 'The Proper Treatment of Connectionism'.

FOOTNOTES

* Parts of this paper were presented at the Annual Meeting of the Linguistic Association of Finland, Helsinki, 13 January 1989. I want to thank Hartmut Haberland and Kristiina Jokinen for useful comments.

1) Quite apart from the fact that Diederich's dictum apparently should be read as an 'isa' statement; cf. his opening phrase: "Konnektionismus ist eine Form der Künstlichen Intelligenz..." (1988:28).

2) Perusing the literature, one often gets the impression that the 'new' cognitive science (including connectionism) is all a North American invention and exploit.
However, as early as 1977, the Finnish researcher Teuvo Kohonen published a book in which connectionist hypotheses are clearly stated and accounted for in mathematical and computational terms. Unfortunately, only recently references to Kohonen's work (now in second printing (1984)) are beginning to emerge in the relevant literature (i.a. Smolensky's (1988) review article).

3) From an ad blurb for a reissue of Minsky & Papert's (1969) book Perceptrons: "...required reading for anyone who wants to understand the connectionist counterrevolution that is going on today." (MIT Press publicity release)

4) 'Distributed' used to be the opposite of 'local' and has to do with the amount of information that is encoded in each single unit of the parallel-processing network. Lately, the distinction seems to have been overtaken by the factual developments: PDP simply is connectionism, and vice versa.

5) A complete set of 200 Wickelfeatures is created separately to characterize phones at word-boundaries (cf. Lachter & Bever 1988: 209-210).

6) To vary an old joke:
Q: What is the mafia's proper treatment of connectionism?
A: Give them an offer they cannot locate.
(But compare footnote 4, above).

REFERENCES

Chomsky, Noam. 1957. Syntactic Structures. The Hague: Mouton.

Chomsky, Noam. 1959. Review of B.F. Skinner, Verbal Behavior. Language 35:26-58.

Diderich, Joachim. 1988. Trends im Konnektionismus. Künstliche Intelligenz 1/88:28-32.

Dietrich, Eric & Chris Fields. 1988. Some assumptions underlying Smolensky's treatment of connectionism. Behavioral and Brain Sciences 11(1):29-31.

Dreyfus, Hubert L. & Stuart E. Dreyfus. 1986. Mind over machine: The power of human intuition and expertise in the era of the computer. Oxford: Blackwell.

Feldman, Jerry A. 1981. A connectionist model of visual memory. In: G.A. Hinton & J.A. Anderson (eds.), Parallel models of associative memory. Hillsdale, NJ: Erlbaum.

Feldman, Jerry A. & David H. Ballard. 1982. Connectionist models and their properties. Cognitive Science 6:205-254.

Fodor, Jerry A. 1975. The language of thought. New York: Crowell.

Fodor, Jerry A. & Zenon W. Pylyshyn. 1988. Connectionism and cognitive architecture: A critical analysis. Cognition 28:3-71.

Hjelmslev, Louis. 1943. Omkring sprogteoriens grundlæggelse. Copenhagen: Munksgaard. [Engl. tr. Prolegomena to a theory of language. Bloomington, IN: Indiana University Press (1954).]

Hoeppner, Wolfgang. 1988. Konnektionismus, Künstliche Intelligenz und Informatik -- Beziehungen und Bedenken. Künstliche Intelligenz 4/88:27-31.

Kohonen, Teuvo. 1988. Self-organization and associative memory. Berlin: Springer [1977].

Kuhn, Thomas. 1962. The structure of scientific revolutions.Chicago, IL: University of Chicago Press.

Lachter, Joel & Thomas G. Bever. 1988. The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. Cognition 28:195 - 247.

Lakatos, Imre. 1970. Falsification and the methodology of scientific research programmes. In: I. Lakatos & A. Musgrave (eds.), Criticism and the growth of knowledge. Cambridge, UK: Cambridge University Press.

Lehnert, Wendy. 1987. Possible implications of connectionism. In: Wilks, ed. 1987.

McClelland, James L., David E. Rumelhart & the PDP Research Group. 1986. Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models. Cambridge, MA: MIT Press.

Minsky, Marvin. 1963. Steps toward artificial intelligence. In: Edward A. Feigenbaum & Jerry A. Feldman, eds., Computers and thought. New York: McGraw-Hill.

Minsky, Marvin & Seymour L. Papert. 1969. Perceptrons. Cambridge, MA: MIT Press.

Pinker, Steven & Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed and processing model of language acquisition. Cognition 28:73-193.

Rees, Nigel. 1983. Graffiti. [Danish tr.]. Copenhagen: Apostrof.

Rumelhart, David E., James McClelland and the PDP Research Group. 1986. Parallel distributed processing: Explorations in the microstructure of processing. Vol. 1: Foundations. Cambridge, MA: MIT Press.

Rumelhart, David E. & James L. McClellan. 1986a. PDP models and eneral issues in cognitive science. In: Rumelhart et al. 1986.

Rumelhart, David E. & James McClelland. 1986b. On learning the past tenses of English verbs. In: McClelland et al. 1986.

Searle, John. 1984. Minds, brains and science. London: BBC (The Reith Lectures; The Listener, Nov.-Dec. 1984).

Shepard, Roger N. 1988. How fully should connectionism be activated? Two sources of excitation and one of inhibition. Behavioral and Brain Sciences 11(1): 52.

Smolensky, Paul. 1988. On the proper treatment of connectionism. Behavioral and Brain Sciences 11(1):1-74.

Wilks, Yorick, ed. 1987. TINLAP-3: Theoretical issues in natural language processing. Las Cruces, NM: New Mexico State University, Computer Research Laboratory.

Winograd, Terry & Fernando Flores. 1986. Understanding computers and cognition: Toward a new foundation of design. Norwood, NJ: Ablex.