

Article

***The Cognizer's Innards:
A Psychological and Philosophical
Perspective on the Development of
Thought***

ANDY CLARK AND ANNETTE KARMILOFF-SMITH

1. Introduction

The sea-slug and the VAX mainframe are both effective processors of information. Yet it is only human beings, and perhaps some higher animals, who are credited with genuine thoughts. Is this mere prejudice, or have we somehow latched on to a genuine joint in the natural order? If it is a genuine joint in nature, what feature or set of features mark it?

The hypothesis to be considered is that there is indeed a joint in the natural order such that humans fall on one side and many other systems (including some quite sophisticated information processors) fall on the other. The joint, we argue, marks a pivotal difference in internal organisation. The representational redescription (RR) model embodies specific hypotheses about the nature of this joint (Karmiloff-Smith, 1979a, 1979b, 1986, 1990, 1992). For the genuine thinkers, we submit, are endowed with

We should like to thank Ned Block, Martin Davies, Julia Grant, Graeme Halford, Hans van de Koot, Mark Johnson, Wendy Lehnert, Dan Lloyd, Jean Mandler, Jay McClelland, Christopher Peacocke, Joseph Perner, and Mike Sharples for comments on earlier drafts, and the University College London Cognitive Science Wednesday Group for their lively discussion of the paper.

Address for correspondence: Andy Clark, Department of Philosophy, Washington University, One Brookings Drive, Campus Box 1073, St. Louis, MO 63130-4899, USA. Annette Karmiloff-Smith, MRC-CDU, 4 Taviton Street, London WC1H 0BT, UK.
Email: andy@twinearth.wustl.edu; annette@cdu.ulc.ac.uk.

an internal organization which is geared to the repeated redescription of its own stored knowledge. This organization is one in which information already stored in an organism's special-purpose responses to the environment is subsequently made available, by the RR process, to serve a much wider variety of ends. Thus knowledge that is initially embedded in special-purpose effective procedures subsequently becomes a data structure available to other parts of the system.

Data from human cognitive development highlight this process. A striking fact about development is that once children achieve proficiency in a given domain, they do not merely repeat the same successful performance. Subtle changes occur subsequently. In some instances such changes can be explained by modifications in the input from the external environment and/or in children's internal resources (e.g. maturation of neural pathways). Similar changes in connectionist networks (e.g. modifications in the input patterns (McClelland and Jenkins, 1990) or recruitment of new hidden units (Schultz, 1991)) cause networks to develop more powerful (i.e. more generalizable) representations. Such representations remain implicit in the network's functioning. Whilst this is the *endpoint* of learning in a connectionist network, in the human case it is the *starting point* for generating redescrptions of implicitly-defined representations. In other words, current connectionist models account rather well for children's *initial* learning in a domain, but they do not yet adequately model the subsequent representational change posited by the RR model (Karmiloff-Smith, 1987).

What theory of learning could encompass the RR concept? In the past decade, the philosophical and psychological literature has been replete with discussions of whether symbolic models (Klahr, Langley & Neches, 1987; Anderson, 1980; Newell, 1990) or connectionist models (McClelland, Rumelhart and the PDP Research Group, 1986) can better account for human development. Such a hard and fast dichotomy is, in our view, premature. In fact human development may involve both types of learning (Bates, Thal & Marchman, 1989; Bever, 1992; Clark, 1989; Johnson-Laird, 1988; Karmiloff-Smith, 1987, 1992; Lachter & Bever, 1988; Norman, 1986; Pinker & Prince, 1988, forthcoming; Shallice, 1988). We will show how the data motivating the RR model helps to pinpoint the need for a move to more complex connectionist models or for hybrid connectionist/symbolic systems. Such, in its broadest outline, is the agenda. We begin (Section 2 below) by discussing a number of connectionist models, the limits of the implicit representations that they generate, and how they fall short of what the RR model calls for. We go on (Section 3) to provide further discussion and empirical support of the increasing cognitive flexibility that the RR process makes possible in human development. Section 4 charts some consequences for connectionist modelling of the idea of a conservative, multi-levelled, redescriptive architecture. In the closing section we explore the wider philosophical implications of the RR model.

2. The Limits of First Order Connectionist Representations

Consider NETtalk. NETtalk (Sejnowski & Rosenberg, 1986) is one of the best known and most impressive examples of what we call a 'first order' connectionist system. The model's goal is to negotiate successfully the problem domain of text-to-speech (grapheme-to-phoneme) transformations. NETtalk takes 7-letter segments of text as input and maps a target window of that input onto an output which codes for phonemes. (The output is sent to a speech synthesizer which converts it into actual sound.) In common with most interesting connectionist systems (*i.e.* any tackling complex, non-toy problems), the net was trained rather than programmed. One of the major achievements of contemporary connectionism is a small set of powerful automatic learning procedures. In the case of NETtalk, a back-propagation method was used. What this means is that the network begins with a random set of weights on its connections, and is given an input. A set of teaching units computes the difference between the network's actual output (in this case a coding for phonemes), which at first will inevitably be wildly inaccurate, and the target output. Automatic procedures then minutely adjust the connectivity strengths to reduce the discrepancy between actual and targeted output, *i.e.* to make it the case that it would do a little better were it given that input again. This is repeated many hundreds of times for many different inputs. After extensive training, the system can be heard (via the speech synthesizer) to babble half-recognizable words. By the end of the training sequence it is accurate and intelligible. An impressive achievement indeed. But now let us raise the question: what does NETtalk actually *know*?

Naturally, NETtalk in no sense *understands* what it is saying; that was not the task. People too can learn roughly how to pronounce Japanese sequences without understanding them. Nonetheless, NETtalk has gone from random babbling to outputs that are systematically related to its inputs. What representations make this possible?

One way to find out is to examine how the network is using its weights to group different inputs. This is commonly done by using various statistical techniques such as cluster analysis (Elman, 1990; Sejnowski & Rosenberg, 1986). Cluster analysis proceeds by:

- (1) giving the network a selection of inputs;
- (2) recording the state of the system's hidden unit activations as a response to each input;
- (3) grouping together the most similar pairs of hidden unit responses;
- (4) examining the inputs to the pair to try to see what they had in common;
- (5) repeating the procedure for pairs of pairs, and so on.

The cluster analysis of NETtalk (Sejnowski & Rosenberg, 1986) revealed similarities of hidden unit response to, for example, p's and b's and,

proceeding up the hierarchy of pairings and pairs of pairings, to all vowels and to all consonants. In some sense, then, NETtalk had learned to respond similarly to all consonant inputs on the one hand and to all vowel inputs on the other. This is quite a striking result. The network has learned to treat a,e,i,o,u, as similar in some respect. So it *knows about vowels*. Or does it? It is true that the network succeeds in a domain which *we* (external theorists) characterize as involving abstractions such as 'vowel' and 'consonant'. There is an important sense, however, in which these abstractions, useful as they are in helping us to see what NETtalk is doing, are not available to NETtalk itself. For as long as knowledge is merely implicit in a task-specific set of weights, it is not possible for networks to communicate that knowledge *directly* to any other network engaged in related problem solving. For example, there is no way in which NETtalk itself or another interconnected network could directly count the number of vowels or consonants it processes. This illustrates the problem of internetwork knowledge transfer: how to exploit the results of successful learning in one network immediately in another, thereby obviating the need for re-training on that aspect of the task. In the absence of such transportable knowledge, networks are unable to progress beyond the limited task-specific use of their knowledge. It is precisely these limitations which, as we shall see in Section 3, the normally developing child subsequently surmounts.

Consider a second example. Suppose that one of the big high street banks has a connectionist network capable of assessing loan-worthiness. The bank has over the years constructed a vast data base about who has or has not proven to be a good risk. However, the data are disunified; no clear patterns emerged that a human operative could spot and rely on. The data are then used to train a connectionist network which learns to assess loan-worthiness with a *significantly higher success rate* than the bank's best human operatives.

Suppose we now attempt to analyse the network. What has it learnt? We might find, by some kind of statistical analysis, that it has learnt to regard *postal addresses* as a useful constraint and has partitioned the space into good addresses and not-so-good ones. For example, it may respond similarly to all addresses in the West End of London. To that extent, it 'knows about' good addresses. That is, it negotiates its problem domain in a way which we might come to understand by introducing the concept of a 'West End address' as a label in a cluster analysis of its behaviour.

Now suppose the loans network is just a small part of a larger system, and the larger system is capable of real communication. The larger system (call it 'bank manager') is asked to advise a leading Australian bank on the matter of loan-worthiness. That is, it is asked for its *theory* of loan worthiness. As things stand, it may well have nothing to say. Much like a human with an automatic skill, it knows how to assess loan applications in a very particular domain (*viz.*—England-1993). But if that is *all* it has learnt, it will have no useful theory to transmit. The set of weights on the

connections of the loan-assessment network won't do, since these represent a solution only to a very *specific* instance of the general problem of loan assessment.

What the system needs, in order to be able to tell the Australian system anything useful, is some more abstract and transportable knowledge concerning relevant factors in loan assessment. Such knowledge might involve general concepts such as 'good postal address', which could be weighed in importance against others such as 'professional employment', 'stable employment record' and so forth. The original network gets by without such explicit abstractions, since it only develops the minimal representations needed to succeed in the version of the task for which it was trained. The point that we wish to stress, then, is that, *left to its own devices*, a network—unlike the child—will not go *beyond* success in the initial task (*i.e.* successful assessment of loan-worthiness in London) to form explicit abstractions which describe the features which it has distinguished in its partitioning of the weight space. It is worth noting that merely training the network to label clusters of the results of its own activity 'good address', 'stable employment' and so on, will not in itself provide transportable representations. This is because they are learnt as a generalization of task-specific exemplars. Thus, consider the related task of loan-assessment in a depressed economy. Here, stability of employment may rate higher than other factors such as salary. The opposite is true in a buoyant economy. Mere provision of label nodes won't help us exploit knowledge gained using data from a buoyant economy to speed up the training of a network that must function in a depressed economy.

Consider another network (which we shall call the Ohm's law network) described in Smolensky (1988). This is a network whose task is qualitative (*i.e.* non-numerical) reasoning about an electric circuit. The network answers questions such as 'If the resistance is increased at location P, what happens to the voltage at R?' It uses a set of feature units which stand for the state of the various parameters (current, voltage and resistance) at particular locations. It also uses a set of 'knowledge atoms', units that encode the various legal combinations of values for the feature units. Between feature units and knowledge atoms there is two-way connectivity. The network embodies Ohm's law ($V = C \times R$) but it does so implicitly. It does not generate transportable symbols for voltage, current or resistance, and it does not explicitly represent the relation between such abstractions. But as always, such implicit representations result in a certain lack of flexibility. To see this, consider the case where there is a systematic change in the task. An example would be where (as an exercise in deviant circuitry, say) a human expert is told to solve problems as if $V = C + R$ (instead of $V = C \times R$). In such cases the lack of explicit abstraction is critical. The symbolic solution, having transportable elements standing for C, V, R and arithmetical relations + and \times , can be systematically modified to reflect the variant task. Not so for distributed representations in a connectionist network. The difficulties arise because connectionist learning tends to

build the relations between entities into the implicit representation of the entity itself. In so doing, it makes it impossible to operate on the relations independently of the entities and vice versa.

In our opinion, the formation of explicit representations provides a system with a kind of flexibility and generality not possible in any first order network. This flexibility and generality manifests itself in a variety of ways, of which the most important seem to be:

- (a) allowing example-independent knowledge of rules (= relaxing a constraint of *statistical* sensitivity which attends first order connectionist solutions);
- (b) coping with systematic changes to the structure of the rules governing a domain;
- (c) intelligent self-debugging without fullscale re-training;
- (d) integration of activities with those of other sub-systems operating on data included in different formats.

Let us expand on these in reverse order. The point about integration of activities is simple but telling. Connectionist systems (of the NETtalk variety) learn a way of negotiating their own particular domain. In effect, they generate a representation scheme that is specifically adapted to that domain. Often, however, it will be necessary for such systems to share their information with one another in order to harmonize their activities. If the information were present not just in its fully distributed form, but *also* in some reduced form which other networks could read and understand, the task of harmonization would become far more tractable.

Intelligent self-debugging presents a related problem. At present, if a network goes wrong it is seldom possible to do anything except re-train it on the corpus of examples. Human self-debugging is often more focused. Thus, to repeat a mundane example used elsewhere (Clark, 1990b), a golfer whose game suddenly deteriorates may isolate the cause of the trouble as the wrist component of her swing. This enables her to take focused action; instead of learning the whole swing anew, she can keep most of it intact and concentrate on the source of the trouble. Such a focus is not possible for a first-order connectionist system in which all its knowledge is inextricably intertwined and no control structures exist capable of unpicking parts of the web while preserving others. This inability to act on part of the learnt representation whilst leaving the rest intact has radical consequences in cases where there is systematic change to the structure of the rules governing a particular domain.

It might be argued that the problems of explicitly representing a network's implicitly represented knowledge are simply an artefact of an early generation of connectionist systems, *viz.* simple pattern associators or 3-layered feedforward networks like NETtalk. This, however, does not seem to be the case. Much progress has been made, it is true, with the general problem of structure-sensitive processing using recurrent networks (*e.g.*

Elman, 1990, 1991; Servan-Schreiber *et al.*, 1988). But structure sensitivity, as presently conceived, continues to involve essentially implicit representations. For example, consider Chalmers' recent use of Pollack's RAAM (Recursive Auto-Associative Memory) encodings to provide a thoroughly connectionist model of direct structure-sensitive transformations (Chalmers, 1990). The RAAM encoding technique (see Pollack, 1990) provides a means of generating compressed connectionist representations of sentence structure trees. Chalmers used the RAAM technique to generate compressed connectionist representations of the structure of a corpus of sentences in active form and a corpus of sentences in passive form.¹ He then trained a second network (which he calls the Transformation Network) to take compressed representations of active sentences as input and to give compressed representations of the corresponding passives as output. These could then be decoded (decompressed) by feeding them back through the originating RAAM device. The transformation net succeeds at its task and generalises to sentences not given in the training phase. Chalmers argues that this shows that it is possible to operate directly on connectionist representations of highly structured items in ways that are required to respect that structure. Nonetheless, this is not tantamount to having an abstract *concept* of the passive or active case. The knowledge remains task specific. It is simply a series of albeit sophisticated compressed representations whose cluster analysis would reveal the implicit representation of the passive or active voice. Again, it could not count the number of passive sentences. More importantly, the same transformation network could not make the *even simpler* transformation of agent/patient reversal in the active voice, *e.g.* transform the string John love Michael into Michael love John. Yet the structural information necessary to perform this task is clearly present in the original RAAM encodings. However, in order to use that information, it would be necessary to train a completely new transformation network. And so on, again and again, for each new variant on the use of the original sentence structure encodings. The transformational networks would learn to access the structural information in only a unidimensional, highly task-specific way. They would not represent the structure as a object susceptible to direct, non-exemplar-driven computational manipulation. This is not the case for the normally developing child. From early on, children redescribe into new representational formats their implicit representations such that they become objects of cognitive manipulation, transportable to other tasks (Karmiloff-Smith, 1979a, 1979b, 1986, and see Section 3, below).

A recent discussion of generalisation in connectionist systems which is sensitive to the kinds of problem discussed above is found in Barnden

¹ Actually, inflections were not modelled. What was modelled was the distinctive word orders of active and passive voice, *e.g.* John love Michael (active); Michael is love by John (passive).

(1991). Amongst a host of interesting observations, Barnden makes two points of special interest to us. First, he recognises the key shortcoming of familiar connectionist generalisation, *viz.* that the generalisations concerned are not available to the system as knowledge items. Instead, the system's knowledge of the generalisation consists in nothing but its ability to pass judgement on new cases in a manner consistent with the judgements it was trained to make on a set of examples. Thus we are asked to consider a network trained to output the conclusion 'atheist' when given as input an example of a communist living in a particular town. Such a network will learn to say 'atheist' for any input involving the feature 'communist'. In a sense, it has then learnt the generalisation that all the communists are atheists. But, as Barnden points out, what the system does not do is to develop a way of constructing an activity pattern that actually represents that generalisation. The generalisation, insofar as it would constitute information for the system, constitutes a piece of knowledge over and above the ability to judge each individual case correctly. In particular, the generalisation should be capable of being accessed and used in the context of other inferences which are in no way concerned with specific individuals but which require knowledge of, for example, the rough pattern of village communists who are also atheists. This information is correctly present in the net imagined and could be accessed by, for example, discovering some resource (*e.g.* across the hidden units) which represents the feature 'communist' and simply locking it on, even in the absence of any specific input concerning individuals. But as Barnden notes, the problem of teaching the network to do this for itself when some inference requires such general information remains unresolved.

The second point of special interest to us concerns what Barnden calls 'anomalous combinations'. The claim here is that familiar connectionist means of generalisation may fail to provide a system with the kind of knowledge needed to cope with, for example, very novel combinations of ideas in new sentences. The idea (clearly related to the kinds of worry made famous by Fodor and Pylyshyn (1988)) is that a net that learns (by training on a corpus) to represent sentences (given as input) as structured, semantically rich items, will be compromised in its ability to represent highly novel conceptual combinations by the inextricability of its knowledge about structure from its knowledge about meaning. Its ability to recognise a given sentence structure will be dependent in part on the particular contents of the items which figure in the structure. According to Barnden, classical approaches are better able to deal with such novel combinations because they separate 'the representation of structure from the representation of components' (Barnden, 1991, p. 17). Hence they can recover the roles of the items even if their combination is semantically anomalous (*e.g.* 'the banana speaks').

Barnden's example again illustrates one of our main themes, which is that it is not always beneficial (though for some tasks it clearly is) to intricately interweave the various aspects of our knowledge about a domain

in a single knowledge structure. Such interweaving poses problems if we want our knowledge to generalise in a flexible way. For the interweaving makes it practically impossible to operate on or otherwise exploit the various dimensions of our knowledge independently of one another. With this in mind, we turn to the RR model and some relevant studies in human cognitive development. These suggest that in many cases human learning begins by forming precisely the kind of highly interwoven knowledge structure distinctive of familiar connectionist models. But, unlike those connectionist systems, the RR model postulates that humans are internally driven to go on to form a series of further representations which allow us to manipulate and exploit our knowledge in increasingly flexible, and mutually independent, ways.

3. The Process of Representational Redescription

The RR concept was developed within the framework of a theory of human cognitive development and tested in experiments on how children acquire language and how they solve problems in physics and spatial tasks (Karmiloff-Smith, 1979a, 1979b, 1984, 1986). Various kinds of developmental data point to three complementary sources of human knowledge:

- (1) It is specified in innate predispositions (at birth and/or via maturation).
- (2) It is acquired through interaction with the physical and the socio-cultural/linguistic environment.
- (3) It involves an internal process of representational change whereby the mind enriches itself from within by re-representing the knowledge that it has already represented.

The third source of knowledge is at the heart of the RR model which posits several levels at which knowledge is represented and re-represented in different formats. Note that the RR model is not a stage model of development *à la* Piaget in which across-the-board changes in representational format would affect cognition in a domain-general fashion. The RR model, by contrast, posits that a non-age-related process of representational redescription occurs *repeatedly* across development, but is highly constrained by the structure of the domain-specific representations over which it operates (see Karmiloff-Smith, 1986, 1992, for a more detailed discussion of these points). Let's briefly take each level of redescription in turn.

The first level is termed 'implicit' or level-I representations. At this level knowledge is represented and activated in response to external stimuli, but it is not available for use by any other part of the system. In other words, it is knowledge *in* the system, but it is not yet knowledge *to* the system. A procedure for producing a particular output is available as a whole to other processes, but its component parts (*i.e.* the knowledge

embedded in the procedure) are not. This is precisely the situation for NETtalk and the other models discussed in Section 2. A further example comes from Elman (1990) whose network also develops implicit knowledge in the sense of 'implicit' as it applies to the RR model. Elman's network attempts to simulate the young toddler's task of learning structure-dependent relations from English-like input strings. The network is fed one word at a time and has to predict the next input state, for example the next word in a string. The difference between this predicted state (the computed output) and the actual subsequent state (the target output) is fed back at every time step. The network exploits a context layer, which is a special subset of inputs that receive no external input but feed the result of previous processing back into the hidden layer. In this way, at time 2, the hidden layer processes both the input of time 2 and, from the context layer, the results of its processing at time 1. And so on recursively. It is in this way that the network captures the sequential nature of the input. Although, of course, the network doesn't manage to predict precisely the right word, it ends up by predicting the right *category* of word (noun versus verb, singular versus plural, transitive versus intransitive verb, agent versus patient, etc.). A principal components analysis of the weight space shows that words falling in the same category tend to line up along the same axis. Thus, when 'boy' or 'book' are used as a subject, they line up together, whereas when 'boy' is used as an object it lines up with other object words in another part of the weight space. In other words, the representations that the network gradually builds in its weight space code for various semantic and syntactic categories. However, like NETtalk, these representations are not transportable: they cannot be isolated and given directly as input to other networks which need to know about linguistic categories for other purposes, for example counting how many plural nouns there are in a string.

This kind of implicit knowledge corresponds only to an early phase in the cognitive development of real children: level-I knowledge. But whereas connectionist models stop at this point, the RR model postulates that human development goes beyond the implicit encoding of knowledge. Once the child has reached consistent behavioural mastery on a particular task, this is the signal for subsequent representation change. In other words, children go beyond success. Via a repeated process of redescription, condensed representations are explicitly defined. This corresponds to level-E1 representations. E1 representations are still not available to conscious access and verbal report. This requires yet other levels of redescription, E2/E3, etc. which we will term 'E+' for convenience in the rest of the paper, denoting levels at which representations become progressively more accessible (for a full discussion of the model, see Karmiloff-Smith, 1986, 1992). Thus development does not merely involve a dichotomy between implicit and explicit linguistically-encoded representations. Rather, it involves multiple levels of redescription, leading to increasing accessibility and flexibility.

Let's take a couple of examples from language acquisition and one from children's drawing to illustrate these different levels. For example, a child learning English will have to acquire knowledge about the determiner system—how to use words like 'a', 'the', 'some', 'my', etc. A similar problem faces children learning French, except that they will also have to learn to differentiate 'un' in its indefinite reference function (equivalent to English 'a') and the same phonological form 'un'² in its numerical function (equivalent to English 'one'). Experimental studies have shown that 3-year-olds successfully express both functions using the correct form (Karmiloff-Smith, 1979a). Let's call this T-1. Just as in the Elman net where 'boy' is represented differently when it functions as a subject or object, likewise in the human case the RR model posits that 'un' is represented differently according to whether the task involves indefinite reference or the numeral. In other words, there are two representations of form-function pairs containing 'un' as the form. But the Elman network is not internally driven to make its linguistic knowledge explicit, *i.e.* in a form which makes it directly useable by other networks. By contrast, the human child goes beyond T-1. A couple of years later in development, T-2, there is a revealing error in children's performance, one suggesting that they have become sensitive to the fact that two functions (indefinite reference 'a' and the numeral 'one') are served by the same surface marker in French. For contexts where the distinction between the two functions needs to be marked, children use 'un X' to mean 'an X' and add a partitive, 'un de X', to mean 'one X'.³ They mark this distinction explicitly in their output. This is later dropped (T-3) such that the single surface homonym is used to mark both functions and the partitive 'de' is then only used as an emphatic.

The RR model explains the passage from T-1 to T-2 as a qualitative change in the nature of the underlying representation. At T-1, two unrelated representations with identical phonological forms serve two different functions. They are not related internally. At T-2, knowledge that was implicit in the child's procedures for understanding and producing the article at T-1 becomes available to the system as data. Then the similarity of the forms is compared internally on the basis of explicit representations. It is now knowledge to the system, rather than just implicit in the system, to continue the distinction drawn above. Having drawn internal links across the explicit representations, the similarity of form is noted and the child marks the functional differences in the external output. At T-2, however, the knowledge is still not available to conscious access or to

² Or the feminine counterpart 'une'.

³ Although the partitive can function as an emphatic in colloquial French, analyses of the results showed that children were not using it in its emphatic function at T-2. They were marking the two different functions. Numerous other lexical items with similar phonological forms but dual functions showed a similar trend (Karmiloff-Smith, 1979a).

verbal report. Children cannot yet explain verbally the relationships and differences between indefinite reference and the numeral, but their new behaviour bears witness to inter-representational links having been formed. The verbal explicitation only becomes possible later still in development via a further level of redescription at E+ levels. Such repeated redescriptions constitute the distinctive power of human cognition to enrich itself from within and ultimately enable metalinguistic awareness. They allow the child to go beyond proficient usage to treat language as an object of cognitive manipulation, something that current connectionist networks do not model.

A second example from language acquisition also demonstrates several levels of representational redescription. This study concerns how children come to understand what a *word* is. Once young children are beyond the very initial stages of language acquisition and are consistently producing both open and closed class words in new, non-formulaic contexts, there can be no question that at some level these are represented internally as *words*, *i.e.* for very young children formal word boundaries are represented and processed as such. But at 3 years of age, children seem to know little if anything *explicitly* about what counts as a *word*. Numerous studies have shown (Berthoud-Papandropoulou, 1978; Bialystok, 1986a, 1986b; E. Clark, 1978; Tunmer *et al.*, 1983) that it is not until about age 6, and for some tasks even later, that children know explicitly that both open class words (*e.g.* 'boy', 'chair', 'silence', 'run', 'think', etc.) and closed class words (*e.g.* 'the', 'any', 'to', 'in', 'when', etc.) are *words*. When asked to count words in a sentence, young children frequently neglect to count the closed class items. When asked directly if 'table' is a word, they agree, but when asked if 'the' is a word, they answer in the negative. Yet at 3 years of age children can perceive, produce and correctly segment words like *the*.

The RR model posits that 3-year-olds' representations of formal word boundaries are in level-I format. By 6 years of age, children's stateable knowledge that *the* counts as a word is, according to the model, in E+ format. So, do we merely need to add to a connectionist network a mechanism that generates linguistic labels for the network's implicit knowledge? The developmental pattern points to a far more complex picture than such a dichotomy would suggest. Indeed, something subtle happens between these two ages. The RR model predicts that there must exist a level of representation between the implicit representations underlying the 3-year-old's correct segmentation of the speech stream into words like *the*, in which formal word boundaries are represented as part of on-line input/output procedures, and the level of representation that enables the 6-year-old's direct off-line metalinguistic reflection about the fact that *the* is a word. This middle level is the E1 representational format. It involves a redescription of information into a format that is accessible to certain tasks outside normal input/output relations but not yet to full metalinguistic explanation.

We set out to test this prediction (Karmiloff-Smith *et al.*, 1991). Previous

studies which asked children whether X is a word or to count the number of words in a sentence, do not engage normal language processing and call for a totally off-line stance. It therefore requires a high level of explicitness (E+ representations). If we are to capture something between the level-I knowledge embedded in totally on-line use of word representations and the E+ metalinguistically accessible knowledge in off-line tasks, we need to devise a way of engaging children's normal language processing whilst getting them to access that knowledge for partially off-line reflection. The following technique did just that job (Karmiloff-Smith *et al.*, 1991, for details).

Three- to seven-year-olds were given a series of partially on-line tasks of a similar design. They listened to a story in which the narrator paused mid-story on open or closed class words. Depending on the task, the child was asked to repeat 'the last word/the last sentence/the last thing' that the storyteller said each time she stopped. No explanation was given as to what counted as a *word/sentence/thing*. This partially on-line technique engages normal language processing and causes an interruption of the construction of a representation of the speech input. Note, however, that the task also has an off-line metalinguistic component (see also Tyler, 1988, for a discussion of the on-line/off-line distinction as it applies to aphasic adults). This calls for explicitly defined representations in at least E1 format. The child must know what the term *word* means and differentiate this from instructions to repeat the last *sentence* or last *thing*. To access and reproduce the last word, the child cannot simply rely on level-I representations. She must focus on an explicitly defined E1 representation of the acoustic input, make a decision as to which segment of it constitutes the last *word* and repeat that segment. In a fourth experiment, a group of subjects' data from the on-line word task was compared with their responses to off-line direct questioning about whether closed and open class items are *words*. For the latter, we simply asked children to help a teddy bear find out what counts as a word and read out one by one a list of words, asking: 'what do you think about *the*, tell Teddy if *the* (*table*, *when*, *silence*, *she*, etc.) is a *word*'. Children were subsequently asked to justify their responses.

On the basis of the RR model it was hypothesized that the off-line task would require E+ representations, whereas the partially on-line task would require only E1 format. It was thus predicted that 3- and 4-year-olds would fail both types of task because their representations of words are still in level-I format, that 5-year-olds would succeed on the partially on-line task by calling on E1 representations, but fail on the fully off-line metalinguistic task, and that by 6 or 7 years they would succeed on both tasks, because by then they have multiple levels of representation with respect to the concept *word*.

These predictions were borne out. A number of the youngest subjects could do neither task well, suggesting that their representations of formal word boundaries were still implicit in level-I format. But our results clearly

established that children as young as 4½–5 years treat both open and closed categories as *words* and differentiate *word* and *sentence*, when the task has an on-line component engaging normal language processing. There was a significant difference, however, between these children's performance on the partially on-line task as compared to the off-line task suggesting that the latter involves E+ representations. In contrast to their performance on the partially on-line task, for the totally off-line task children rejected several exemplars from the closed class category as *words*. Only the 6-year-old subjects and older were equally successful at both tasks.

In sum, despite lengthy modelling and help from the experimenter, very young children simply could not do the partially on-line task. Yet their fluent language and lack of segmentation errors suggest that they do represent formal word boundaries for the majority of words they use and understand. But at this level, they were not yet ready to go beyond behavioural mastery. There were other children around 4½ years old who could not do the task immediately but who, with one-off modelling for a few open class words only, were able immediately to extend the notion of 'word' to *all* open and closed class categories. Their level-I representations were ready for redescription into E1 format, generated from outside. However, just slightly older 5-year-olds had spontaneously undergone the redescriptive process themselves. They showed immediate success even on the practice story. In other words, their earlier implicit representations sustaining fluent output had become data manipulable for other tasks. But they were not as successful on the off-line task. Finally, the oldest subjects' representations showed signs of having undergone further redescription into E+ formats, because they were able to consciously access their knowledge and to provide verbal explanations as to what counts as a 'word' and why. This process of multiple redescription of knowledge that becomes increasingly accessible to different tasks is precisely what seems to be missing from connectionist models of learning.

The foregoing examples from language development make representational redescription look like a rather simple phenomenon. But as soon as one considers a more complex behaviour, specific constraints on representational redescription can be more clearly pinpointed. A number of studies of children's drawing were used to explore these more general constraints on representational change (Karmiloff-Smith, 1979b, 1984, 1990). Let us take a particularly simple yet revealing task involving children's drawing.

Children between 4 and 11 years old each produced six drawings. They were first asked to draw a house, then a house that does not exist. The same instructions were given for drawing a man and an animal but, to keep the discussion brief, we will only deal here with man-drawing. The rationale for the experiment was as follows: by roughly 4½ years of age, children can rapidly draw a man. Their drawing procedure seems to involve a fluent set of sequenced instructions of the type: 'draw head/

draw body/draw limbs'. This contrasts with their ability to draw less basic objects (e.g. a sellotape holder—see Van Sommers, 1984) for which output is neither rapid nor formulaic (see Freeman, 1980, 1987, for discussion). Asking children to draw a man that doesn't exist should force them into operating on the components of the representations underlying their drawing ability and into deviating from the normal sequence of operations.

Full details of the study can be found elsewhere (Karmiloff-Smith, 1990). It was hypothesized that children would first be unable to isolate and further manipulate any of the components of the representations that allowed them to successfully produce normal drawings. This hypothesis was borne out by the results of some of the youngest children. Although they announced that they were going to draw a pretend man, they proceeded to use their usual sequence of operations to produce a normal man. Their man-drawing procedure was generated by only level-I representations; they could not manipulate the component parts. However, other children did achieve flexibility. What was of particular interest were the constraints on the types of change children introduced.

Striking developmental differences emerged between the 4–6-year-old age group and the 8–10-year-old age group. These differences are illustrated in Figure 1. The modifications made by the younger children involved changes in the values of variables of size and shape and also involved certain types of deletion (see (a), (b), and (c) in Figure 1). Older children's modifications were different in nature. They inserted extra elements from the same conceptual category, changed the position and/or orientation of the elements or the whole, and inserted elements from different conceptual categories (see (e), (f), and (g)). These involve inserting sub-routines, re-ordering of sequences, interrupting and treating components as separate parts, and using representations from other conceptual categories. Note that although 4–6-year-olds' value reassignment often involves changes midprocedure, it does *not* involve interruption or re-ordering of the sequential operations. Note also the difference between the two age groups in deletions (see (c) and (d)). In the younger case, deletions tend to appear at the end of the drawing procedures; in the older subjects they involve interruption, insertion of a sub-routine, and continuation until the drawing is finished.

These differences demonstrate the older children's increasing capacity for interrepresentational flexibility and integration of knowledge across domains. It is this kind of increasing flexibility, the capacity to focus on component parts of knowledge, and the relaxation of sequential constraints, the ability to re-represent knowledge implicit in an efficient procedure and to use it subsequently as manipulable data, which is especially hard to achieve using only connectionist resources.

But is such flexibility merely the result of external pressures? The RR model postulates that external stimuli, for example new input, failure of a procedure to bring about a certain goal, communicative pressures from others, etc., are not a necessary condition for representational redescription

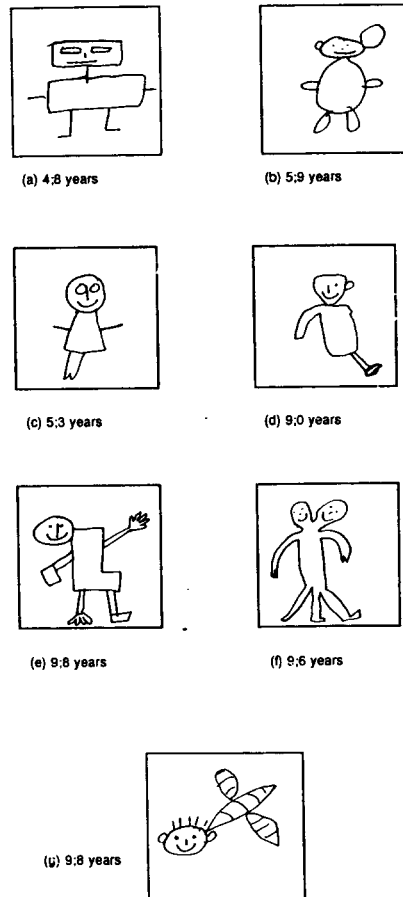


Figure 1 Children's productions (after Karmiloff-Smith, 1990)

to take place. In the main the RR process is generated spontaneously from within. It does not depend on information currently coming into the system for processing.

The RR model is, in this sense, consistent with data indicating U-shaped learning curves (Strauss, 1982) and explains them as a direct result of changes in internal representation of a problem domain: changes that cause new failures as the (temporary) price of a more efficient overall style of representation. It is, of course, an open and empirical question to what extent (and at what level of granularity) such U-curves actually occur. Thus initial criticisms of the connectionist model of the past tense suggested

that that model was vitiated by a reliance on a change in *external* inputs to force the U-shaped learning curve. Such external forcing was not a plausible account (it was claimed) of the origin of such a behavioural profile in children (see *e.g.* Pinker & Prince, 1988). It has recently been shown, however, that the actual behavioural phenomena are much more subtle than they at first appeared (involving 'micro' U-curves for specific verbs at various developmental moments) and that a more complex connectionist model fits this fine-grained data quite well (see Plunkett & Marchman, 1991, 1993). Nothing in the RR model, however, *depends* on U-curved data (whether macro- or micro- in style). Indeed, whereas some physics tasks used within the RR framework did show clearcut macrodevelopmental U-shaped learning curves (Karmiloff-Smith, 1984), other psycholinguistic tasks on the use of cohesive markers in discourse did not (Karmiloff-Smith, 1985). In both cases, the RR model invoked internal representational change, but only in one case did this reveal itself in the behavioural data. What is essential to the RR process is that changes in the mode of internal representation can occur even after surface success is achieved, and can be generated by internal rather than solely external pressures. Such changes need not (though they may) result in any temporary spate of new failures. The key point, then, is that although the RR process is indeed *sometimes* generated by externally provoked stimuli, it is a predominantly endogenous process hypothesized to differentiate human cognition from that of most (all?) other species (Karmiloff-Smith 1979a, 1979b). Classical empiricism would predict that no change should occur in the case of consistent success, *i.e.* positive feedback. Yet the examples discussed in this section show that children frequently introduce change after they are already successful (Brown, 1990; Brown, Bransford, Ferrara & Campione, 1983; Karmiloff-Smith, 1979a, 1979b, 1984).

The RR model, then, is not a data-driven or failure-driven model of change. Although negative feedback is often involved in change, positive feedback also plays an important role in generating internal representational change. Development involves cycles of different phases in which either the external or the internal environment plays a more predominant role at different times. While connectionist models are superbly suited to responding to externally-driven change, it is unclear how they would model the RR process of self-generated internal change. The evidence from child development suggests that there are some real benefits to be obtained not only from having implicit representations which allow a system to perform rapidly and efficiently, but also by developing explicit representations which allow a system to become more manipulable and flexible.

4. The Way Ahead

We are now in a position to define the class of first order connectionist models. This class is essentially the class dealt with in Smolensky (1988) and is characterised by three features:

- (1) The network's learning is purely *example-driven*. Any change in internal representation reflects a change in the statistical properties of the input-output pairings used in training.
- (2) Knowledge of rules in such networks is always *emergent*. They do not depend on symbolic expressions that stand for the elements of a rule. Instead they exploit a multitude of subsymbolic representations whose complex interaction produces behaviour which, in central cases, fits the rule.
- (3) They have no self-generated means of analysing their own activity so as to form symbolic representations of their own processing. Their knowledge of rules always *remains* implicit unless an external theorist intervenes.

But not all is negative. The moral is that first-order connectionist networks (*i.e.* ones like NETtalk) are a nice solution to problems of a certain type, *viz.* ones in which (a) the problem domain involves the subtle weighing up of many small factors, and (b) the problem domain is relatively *stable*, that is we are not required to deal with systematic shifts in the relative weighing of factors or expecting to have to cope with any very novel situations. In such cases it is possible to get by with representations that severely limit the organism's flexibility and adaptability. In fact, we may *embrace* such limitations as the price we pay for speed and fluency. Likewise for the implicit representations sustaining children's fluent language production.

Nonetheless, this style of representation is not always sufficient. No system in which rules are always merely implicit and emergent can, in our view, exhibit the kinds of higher order flexibility and creativity found in humans. Only explicit rules have the genuine, systematically manipulable components that make radical flexibility possible. Our argument is thus meant to do for (some of our) knowledge of rules what Fodor and Pylyshyn (1988) attempt to do for the data structures upon which the rules operate. It is only by having representations (rules, theories, etc.) that are (a) explicit and (b) can thus *themselves* stand as data structures to be operated upon, that a system can fluently adapt its rule-following behaviour to changes in a particular domain.

We have seen that explicit representations are a pervasive and central feature of creative and flexible human cognition and that they are lacking from standard connectionist models. How should the connectionist respond? As we see it, there are four options:

- (1) Give up connectionism entirely and revert to a thoroughly classical approach (despair).
- (2) Augment connectionist-style networks with the symbol structures of natural language (a representational leap).
- (3) Combine elements of connectionism and classicism in a single system (hybridization).

- (4) Use thoroughly connectionist resources in increasingly sophisticated ways (more of the same).

Of the options, we shall not consider giving up connectionism altogether. With others, we believe that connectionist approaches definitely have something important to offer the task of understanding human cognition (Bates & Elman, 1993; Clark, 1989; Karmiloff-Smith, 1987, 1992a, 1992b; McClelland, 1989). They may provide a formal mechanism to account for Piaget's intuition about the compensatory processes of assimilation and accommodation. The question is how much can connectionist approaches offer? We shall first briefly deal with option 2 and then focus on options 3 and 4 ('hybridization' and 'more of the same').

The second possibility deserves a brief mention, although it runs against the ideas embodied in the RR model (Karmiloff-Smith, 1986). This second solution involves arguing (Clark, 1989, Chapter 7; Smolensky, 1988; Dennett, 1991) that the human mind deploys essentially connectionist style representations but augments itself with the symbol structures of natural language in the public domain. On such a view, the expert banker's ability to adapt rapidly to the radically changed domain would be explained by supposing that she is in possession of a set of linguistically formulated rules which she can (naturally) re-configure as the need arises. But we saw that the developmental sequence is *not* from lack of flexibility on to linguistic expression and thence to full flexibility. Instead, children often show signs of representational redescription and exhibit flexibility *before* they are able to give linguistic expression to their knowledge. Moreover, developmental data suggest that frequently the ability to re-configure the linguistic description of a task is not of any help. Theories that make the distinctive cognitive characteristics of humans dependent on an ability with public language seem, in general, to get the cart before the horse. It is more plausible to see our abilities with language as one *effect* or *product* of a deeper underlying difference in the redescriptive architecture of our cognitive apparatus, a difference that may group us with some non-linguistic higher mammals but separate us from hamsters, sea-slugs and standard connectionist networks (Karmiloff-Smith, 1979a, 1979b, 1986; for a more recent discussion of these ideas, see Clark, 1990a, 1990b). Language itself, then, emerges as just one more level of redescription, albeit one that provides rich manipulability and a powerful means of cultural transmission. Note that the RR model does not claim that *all* symbolic representation reduces to redescription of non-symbolic processes. For example, some knowledge may be directly stored in linguistic form, and some procedural knowledge is never available to symbolic recoding, even in adults (Karmiloff-Smith *et al.*, 1993).

The basic point is just that the careful study of developmental data shows that the increased behavioural flexibility often pre-dates the subject's ability to formulate or understand linguistically encoded expressions about the components and rules governing the domain. The possibility

of such expressions is an effect of a prior process of representational redescription, not a substitute for it.

Let's go on to options 3 and 4. A relatively straightforward case of hybridization is exemplified by the PRO system described in Lehnert (1987a, 1987b). PRO is a system that learns to pronounce English words. It negotiates in essentially the same domain as NETtalk but, unlike NETtalk, PRO uses classical heuristic methods to build complex structures in memory and then operates on those structures using broadly connectionist search techniques, *viz.* spreading activation and numerical relaxation. This combination allows PRO to use classical case-based reasoning abilities (actually PRO remembers only fragments of actual training items which are then further organized into mutually excitatory and inhibitory groupings) whilst resolving conflict between potentially relevant cases in the fluent highly parallel mode distinctive of connectionist processing. PRO thus uses connectionist techniques to buy speed and fluent conflict resolution whilst availing itself of many of the more structural resources of classical AI. The model pays a price for these structural resources insofar as it does not exploit the ability of many first-order connectionist systems to generalize directly in virtue of the superpositional storage of distributed representations. The ultimate potential of such hybrids will require a thorough assessment of such cost/benefit tradeoffs.

A somewhat different approach is taken by Touretsky's BoltzCONS (Touretsky, 1989). BoltzCONS is a fully fledged connectionist symbol processor insofar as it can generate and act upon complex symbol structures. It attempts to tackle the issue of representing structure by allowing the recombination of representations into complex expressions. It has the ability to refer to its own data-structures without recapitulating a full representation of their content (what Newell, 1980, called 'distal access'). Yet it is not just an implementation of a classical system, for it avails itself of powerful connectionist search techniques including free content addressable memory, partial pattern matching, and the efficient exploitation of multiple cues. It does not, however, use distributed, subsymbolic representations, although such an extension is deemed 'easily imaginable' (Newell, 1980, p. 28).

At this point it is important to note that the RR model raises two potentially distinct computational issues. The first is: how does redescription occur? That is, what *kind* of internally driven computational process yields the higher level redescriptions? The second is, in what does the redescription consist? That is, what is the character of the higher level representational formats? There is thus a question about *process* and a question about *product*. And it is at least conceivable that these might cross-classify the connectionist/classicist issue. Thus we might have a connectionist developmental process whose product is a symbolic representational language. Or we might have a more classical process which finds the kind of input/output representations that make a task tractable for a connectionist system.

With this distinction in mind, we can briefly consider two proposals which fall into our fourth ('more of the same') category. The first focuses on the *product* of redescription. The idea here is to design a thoroughly connectionist model which, while exploiting the special properties of connectionist representation (partial pattern matching, content addressable retrieval, graceful degradation, speed of processing, example-driven category and concept formation) is nonetheless able to form structured expressions whose parts can be operated on by other computational processes.⁴ Hints of such systems have been around for some time, for example attempts to formulate a connectionist solution to the variable binding problem (see Smolensky, 1987; Touretsky & Hinton, 1986). There are proposals to equip connectionist systems with two kinds of representation, one of which is a condensed, manipulable version of a larger (expanded, microfeatural, distributed) representation (Hinton, 1988; Pollack, 1990). The condensed representation (formed by the iterative compression of a hidden unit activation pattern) could then be used as a kind of stand-in where the fully expanded version is uneconomical or difficult to use. The use of such condensed representations may be a step towards solving the problem of effective inter-network communication. It should be noted, however, that it is only by extensive training that current systems are able to exploit the information contained in such representations. We saw, for example, that a network can be trained to perform direct structure-sensitive transformations (active to passive sentences and vice versa) on such compressed representations (Chalmers, 1990). However, if we wish to exploit the sentence structure-dependent information contained in the compressed representations for some new purpose or task, this would involve the complete retraining of a new transformation network. Such networks never become 'grammarians', whereas (to some extent) children spontaneously do.

Mechanisms capable of helping to explain the *process* of redescription are harder to find. One example is Mozer and Smolensky's technique for the 'skeletonization' of networks. The idea is to use an automatic procedure which takes a successful 'trained-up' network and computes a 'measure of relevance that identifies which input or hidden units are most critical to performance', and then automatically deletes the least relevant units. It turns out (Mozer and Smolensky, 1989, pp. 4-5) that this can improve the

⁴ One of us (A.K-S) at present on sabbatical at Carnegie Mellon University, is currently working on two such projects. The first is an attempt to model data from empirical work on adults' and children's explanations of different classes of speech repair (Karmiloff-Smith *et al.*, 1991) and to explore the internal representations of language learning networks to ascertain differences in those classes of representation that do become available to metalinguistic explanation and those that do not. The second model is being explored with Jonathen Cohen and Randy O'Reilly, in an endeavour to capture how a concept such as ordinality could emerge as an explicit, transportable representation across different networks performing mathematical and non-mathematical tasks that all involve ordinality implicitly.

system's ability to generalise, speed up learning, and yields a network whose behaviour can be understood 'in terms of a small number of rules instead of an enormous number of parameters'. There is, of course, a price to pay for this; the reduced size network can no longer deal with every nuance of the training set (*ibid.*, p. 15).

But if we combine the idea of skeletonization with the RR picture of a process of redescription (in which the original representations are preserved), instead of replacement (in which they would be overwritten), a powerful prospect opens up. If skeletonization were thus conservative, then the original (untrimmed) network would be preserved—what would be trimmed (*i.e.* operated on by a process of redescription) would be a duplicate, and not (as it is in the Mozer and Smolensky model) the original network. That would bring the network closer to the possibilities available to children, *i.e.* make it possible for the system to deploy the fully nuance-sensitive network when the need arose, but to avail itself of the skeletonized one to yield more powerful generalisations, etc. For imagine a connectionist system with (a) an inbuilt skeletonization procedure and (b) the copying abilities necessary for conservative skeletonization. Such a system might be capable of rapidly adapting itself to a new but systematically (relative to the rules) altered domain by first copying a highly skeletonized network, which was adapted to the original domain, and using that to do the learning needed for the new domain. Such a network will in effect be pre-adapted for the new domain. Its architecture of layers and units, and some of the weightings, may be such as to give it a serious head start over a network with random starting weights. The upshot would be a connectionist way of explaining the phenomenon of 'transfer of learning' to a systematically altered but related domain.

What is attractive about the skeletonization procedure is that it could provide a *connectionist* version of the idea of internally driven developmental change of a kind that enables a system to go *beyond behavioural mastery* in the way envisaged by the RR model. Thus, whereas a standard network will cease to alter its weights once it is successful at negotiating a target domain, a system with a built-in skeletonization procedure could take such on-line success as the signal to engage in a further process of self-analysis and reorganization. Moreover, this process has the nice property of beginning to loosen the system's ties to the total statistical profile of the training set. For example, in one of the cases discussed by Mozer and Smolensky, a network that originally embodied a sensitivity to both a number of partial cues (statistically salient in the training data) and a number of sufficient cues, was able to purge itself of all but the sensitivity to a single sufficient cue. Skeletonization thus provides one example of a connectionist technique by which a system could come to work on (and form increasingly abstract versions of) its own representations without need of further pressure from exogenous sources. Such self-redescription and self-reorganization is at the heart of the RR model.

We therefore conjecture that if connectionist systems are to model human development adequately, they must be able to:

- (a) treat their own representations as objects for further manipulation;
- (b) do so independently of prompting by continued training inputs;
- (c) retain copies of the original networks; and
- (d) form new structured representations of their own knowledge which can be manipulated, recombined and accessed by other computational processes.

A potentially promising ending for our survey of possible options is provided by some recent exploratory work (borderline between options 3 and 4) by Finch and Chater (1991). Finch and Chater describe a system in which a connectionist network isolates salient statistical regularities which are then transformed into explicit symbolic representations of the abstract structure of the inputs.

In a little more detail, the task of the system is to learn about linguistic structure on the basis of strings of unlabelled linguistic input, much like the Elman network described earlier (Elman, 1990; see also Braine, 1987, and Maratsos & Chalkley, 1980, for a discussion of distributional analysis theories of child language acquisition). The network seeks statistical correlations between words (appearing as near neighbours) in a large training corpus, and uses this information to induce syntactic categories such as verb, adverb, adjective, etc. Unlike the Elman network, it does not yet deal with long-distance structure-dependent relations. However, it does address the issue of explicit representations head-on.

The Finch and Chater network proceeds by first learning so-called 'bigram statistics' for the corpus, *viz.* which pairs of words are found as neighbours in sentences in the corpus and with what frequency. This information, which is very naturally acquired by connectionist means, is then transformed into a tree-like structure which taxonomizes the lexical items in a standard symbolic form, *i.e.* as consisting of groups of words that share the role of proper names, of quantifiers, of adjectives, etc. It is, indeed, no surprise that mere statistical analysis can yield such categories. Where Finch and Chater's approach differs from previous ones is in using a neural network to directly compute the bigram statistics of the input corpus and then deliberately setting out to use this information to produce (via cluster analysis) a fully symbolic description. The difference, then, is that (Finch and Chater, 1991, p. 18):

While Elman sees cluster analysis simply as a means of analysing network behaviour, we see it as an integral part of our model . . . Cluster analysis plays the role of the interface between a vector space representation in the network and a structured symbolic representation of the data.

Thus expressed, the Finch and Chater project addresses the central issues with which we have been concerned. For the key move is to somehow make the information reflected in a set of learnt weights—information

about regularities in the input data—available as a distinct computational object for the overall system. It is not at all clear, however, to what extent Finch and Chater's system realises this goal by spontaneous, internally-driven activity in the system, and to what extent further human intervention is involved in the final derivation of a symbolic structured representation. They claim only that a component of this transition is handled by the network's automatic gathering of bigram statistics. The final cluster analysis is still undertaken from 'outside'. The basic spirit of the approach is, however, encouraging. For it recognises the explicit symbolic representation of a network's achieved state of knowledge as an important research goal. Finch and Chater do not go on to discuss what specific benefits may accrue to a system in virtue of the formation of such representations. Our claim is that the formation of such representations is necessary to enable the system to distinguish between the abstract structure of a domain and the particular contents invoked in the training cases, and that such a distinction is vital if the system is to apply its structural knowledge to very novel cases or to deploy it in a domain that requires the systematic transformation of the broad strategies developed to cope with the original cases.

5. The Philosophical Status of the Innards

There is a difficult philosophical problem lurking in the background of our discussion. It concerns the *kind* of importance that should be attached to a being's having innards of the computational kind we describe. Let us call any being whose internal economy supports a process of representational redescription a Redescriber. We can then ask:

Is being a Redescriber a conceptually necessary condition of being a true believer?

Notice that nothing in Sections 1–4 commits us either way. It is consistent with those observations to hold either that humans are Redescribers but that some other thinking beings (say, Martians) will not be, or to hold that as a matter of empirical (but not philosophical) fact, all thinkers are Redescribers.

Notice, also, that the question is already (albeit in a slightly less specific form) the topic of some lively debate. Thus the mythological version of Daniel Dennett (the real thing being more complex though, we suspect, perhaps somewhat less consistent) is famous for claiming that *innards* (and, a fortiori, redescriptive innards) *don't count*. Thus we read in a flagship paper (Dennett, 1987, p. 29; emphasis in original):

All there is to being a true believer is being a system whose behavior is reliably predictable via the intentional strategy.

What this means, in Dennett's model, is that if the ascription of beliefs, desires and so on gives you a powerful enough predictive toe-hold on a creature's behaviour, then it counts as a genuine believer or cognizer, as a haver of thoughts 'in the fullest sense of the word'. Nonetheless (and here we have a clear symptom of the delicacy of these issues) the same writer feels compelled to admit (Dennett, 1988, pp. 542-3):

If one gets confirmation of a much too simple mechanical explanation, this really does disconfirm the fancy intentional level account.

Dennett's 'anti-simplicity' intuition strikes us as strange. There is no obvious *philosophical* reason why simplicity should matter. Yet it is an intuition shared, in various forms, by many contemporary writers. Such intuitions are rooted in the consideration of cases in which apparently intelligent (and voluminous) behaviour is produced by maverick means. For example, we can imagine a super-fast, well-programmed, giant look-up table which, by encoding a canned response for every possible input, yields a magnificent simulacrum of human behaviour. In this case we are loath, it seems, to concede that the being facing us is a genuine believer. But why? What constraints, if any, operate against the look-up being?

It is plausible to suppose that a true believer must grasp the concepts that structure the thought ascriptions that are true of her. Thus, if a being is to count as having the thought that Mary is a successful neuroscientist, that being needs to grasp the general concepts of, for example, 'success' and of 'someone being a neuroscientist'. But what does that involve?

Evans (1982) suggests that we find a full grasp of a concept only where we find a subject whose cognitive resources display a particular kind of interanimation. The claim (known to philosophers as the 'generality constraint') is that if we are to be genuinely warranted in crediting an agent with the thought that *a* is *F*, then (for every predicate *G* of which that agent has a conception) we *must* be satisfied that the agent has the ability to think *a* is *G* (see e.g. Evans, 1982, p. 104). More generally, if someone is to be credited with the structured thoughts '*a* is *F*' and '*b* is *G*' (e.g. 'Mary is hot' and 'John is cold') they must be capable of the thoughts '*b* is *F*' and '*a* is *G*' (e.g. 'John is hot' and 'Mary is cold'). The argument for the necessity of such interanimation of contents (if we are to find the structured thoughts in question) is that to count as grasping the *structured* thought 'John is hot' we must grasp its component concepts (e.g. 'is hot'). And to grasp such a concept is to understand that it picks out a property with a certain width of application. Thus you must know that many items (not just John) are capable of being hot.

Of course, we need not (and indeed Evans did not) demand that *every* concept be fully capable of figuring in the concept slot of every other thought we can have. But the focus on widespread interanimation of

cognitive resources as a requirement of conceptualized thought strikes us as entirely correct.

Let us agree, then, that beings count as grasping a general concept only insofar as they are capable of deploying that concept in a host of contexts involving their other cognitive contents. This notion of conceptualized thought is thus *not* of the all-or-nothing variety. Creatures count as having more or less conceptualized thought according to the degree of freedom of their application of stored information. How might such a constraint link up with the idea of redescription?

Recall once again the point and strategy of the RR process. The point is to enable an organism to exploit fully information that is already encoded in various special-purpose ways. The strategy is to redescribe progressively the information in ways that relax the constraints on its use. Here, then, we have an empirically-rooted parallel to the requirement of conceptual interanimacy. For interanimation *demand*s that a high-level thinker (a haver of conceptualized thoughts) possess a cognitive architecture in which there are increasingly fewer barriers to the use of a given concept in relation to other conceptualized information the organism possesses. Representational redescription creates new informational structures which by-pass the barriers characteristic of their lower level cousins. It is no surprise, then, that the class of beings who seem to meet Evans' constraint are also the class of beings who engage most relentlessly (according to the RR model) in the process of representational redescription. Real thinkers, we argue, meet the Generality Constraint by courtesy of the endogenous drive towards redescription which increase the general availability of the information they are privy to. Nonetheless, the fact that Redescribers will exhibit the required degree of flexibility does not establish redescription as a philosophically necessary condition of the grasp of concepts. Rather, we may cast redescription as explaining how it might be that humans come to exhibit the behavioural flexibility they *must* exhibit (under normal conditions) to count as true believers.

The natural response to such a behavioural reading of the flexibility requirement is to raise once more the spectre of the Giant Look-up Being. Its behaviour (by hypothesis) meets the requirements of *surface* flexibility. But surely, the intuition goes, it *cannot* count as a true believer.

It is tempting to respond by elevating the RR process into a conceptually necessary condition of concept possession. That is, it is tempting to unpack the anti-look-up intuitions as arguing that *only Redescribers* will count as true believers. This, however, would be an over-reaction.

Here is an alternative analysis which seems to us to do justice to the anti-behavioural intuitions and to give the proper force to the redescription hypothesis. The story, which we cannot fruitfully elaborate (if we had a good empirical account of consciousness—if anyone did!—we'd be writing a very different paper) goes like this. Suppose it were the case that the kind of computational architecture deployed by the look-up being is not such as could give rise to any kind of *conscious experience*. Suppose, that

is, that once cognitive science has a model of the computational roots of consciousness, it becomes *obvious* that a look-up architecture will not support such a phenomenon. Relative to such a model of consciousness we may (let us imagine) spot the Redescribers as one class of beings whose innards *are* of the (scientific) kind needed to support such experience. Consciousness, on such a model, would be a side effect of the drive to form ever more abstract and structural representations of one's own stored knowledge. (This link between multiple levels of redescription and consciousness has always been at the heart of the RR model.) Now suppose that (as seems likely) it is a *philosophically* necessary condition of being a true believer that you can be capable of consciously experiencing the structured contents of (at least some of) your thoughts. Then we can diagnose the Anti-Look-up-Being intuitions as a distorted vision of a possible empirical fact—but one posing as a genuinely philosophical constraint on the class of concept-users. The intuition (purely empirical) is that the look-up being would have *no mental life*, no conscious experience whatsoever. The philosophical demand that a genuine believer be capable of experiencing the content of (some of) her thoughts then presents itself (in disguise) as the idea that it is the look-up architecture that itself *directly* undermines the status of the being as a candidate believer.

If we are right, consciousness and rich behaviour collectively exhaust the conceptual constraints on the class of beings capable of structured thought. The role of redescription is, then, twofold. It is the means whereby we achieve higher orders of flexibility and (more speculatively) the root of our conscious awareness of structured mental states. It follows that (philosophically) you don't need to be a Redescriber to be a believer, but (empirically) it may turn out to be pivotal.

6. Conclusions

We have tried to keep the main focus of this paper on what we see as the basic psychological and computational issue, *viz.* how to achieve flexibility, manipulability and transportability within a broadly connectionist setting.

Less ambitiously, we hope that at a minimum, the problems we have raised display the urgent need to devise learning systems that move *beyond* the acquisition of merely implicit knowledge of structures, rules and abstract features. We believe that the developmental studies provide clear evidence that such a progression occurs in human cognition and that it involves a (conservative) series of phases in which constraints on the use of acquired knowledge are gradually relaxed. The crucial question—by what concrete computational mechanisms may such relaxation be achieved?—remains unresolved. What stands firm is the importance of such a mechanism in human cognition, and the problems it raises for familiar connectionist models of representational change.

There is, we believe, a very general moral to all this. It is that the future of connectionist modelling may lie not (as so often at present) in using connectionist learning simply to densely interweave a wide variety of kinds of information (syntactic, semantic, etc.) in a single knowledge base, but instead in using it *also* to unearth, by statistical means, various features and properties which may then be separated out and treated as distinct and independent parameters for future manipulation, just as in the case of human development. Despite some promising hints, then, it seems clear that no current connectionist-inspired models provide exactly the resources the RR model of human development has isolated as essential.

We believe that the conservative nature of the multiple layers of representation posited by the RR model brings additional benefits. For the interweaving of the knowledge about various features (*e.g.* semantic and syntactic) that characterise a problem domain buys a certain on-line fluency and provides a means of supporting easy interactions amongst such features. These interactions are surely as important to high level cognition as the more tightly focused manipulative abilities displayed in more metacognitive tasks. The RR image (in which the densely interwoven knowledge base is preserved but layers of representation are added that strip and separate out various dimensions and feature-types) allows us to have our (interwoven) cake and eat it.

Finally, we note that systems capable of providing these benefits will probably need to resemble the RR model even more closely than the hybrid systems described above in one further respect. They will probably need to generate multiple levels of increasingly abstract and manipulable representations of the basic knowledge they acquire by connectionist means. To try to move from statistically-based, fully interwoven connectionist representations directly to a single highly abstracted, symbolic form is a representational leap and may turn out to be an intractable task. Current systems are prone to fall back on human intervention to bridge this gap. But since the goal is a system that is self-driven to automatically generate the more symbolic representational forms, such interventions cannot be tolerated for long. Once the human is expunged from the model, it may become apparent that the transition needs to be made by degrees. The idea is nicely presented in a recent introduction to a special issue of an A.I. journal devoted to hybrid systems research. The guest editor (Thornton, 1991, p. 7) asks:

Why . . . do hybridists tend to concentrate on the one level model? Putting it another way, why do they typically attempt to obtain symbolist properties directly in a connectionist model or vice versa? It seems to me that the most plausible hybrid arrangement could be one involving a hierarchy of virtual machines with connectionist flavoured machines at the bottom and symbolist flavoured machines at the top end.

This vision is at the heart of the RR model also (Karmiloff-Smith, 1990, p. 79):

In my view, situating sequential constraints within this broader context of multiple representational levels, and linking redescription processes to cognitive flexibility and ultimate conscious access, offers a new and deeper account of developmental change.

We have sketched a particular picture of the cognizer's innards. It is a picture in which the true cognizer is a multi-faceted representor of its (external and internal) world. Such a being combines the basic behavioural fluency made possible by, for example, first-order connectionist modes of representation, with the kind of higher-order representations needed for flexible use of that information. This latter feature was argued to require something more than the relatively unstructured resources of a first-order connectionist system. Whether this amounts strictly to the need for classical symbol manipulation or only to a need for fancier forms of connectionist representation must remain an open question. What seems certain is that the genuine cognizer must somehow manage a symbiosis of different modes of representation—the first-order connectionist and the multiple levels of the more structured kinds. The RR model was seen to guarantee such a symbiosis by invoking a developmental process in which the more structured representations arise as a result of the system's endogenous drive towards the analysis and re-representation of its own cognitive states. Humans, unlike sea-slugs and NETtalk, are blessed with redescriptive and self-analysing innards. Such innards build structured thought on the cognitive bed-rock of fluent but limited (because special-purpose) problem-solving procedures. Both are of the essence.

*Department of Philosophy
Washington University
St. Louis MO 63130-4899
USA*

*Cognitive Development Unit
Medical Research Council
4 Taviton Street
London WC1H 0BT*

and

*Department of Psychology
University College London
UK*

References

- Anderson, J.R. 1980: *Cognitive Psychology and its Implications*. New York: W.H. Freeman.
- Barnden, J. 1991: Connectionism, Generalization and the Propositional Attitudes: A Catalogue of Challenging Issues. New Mexico State University Research Report MCCS 91 211.
- Bates, E. and Elman, J.L. 1993: Connectionism and the Study of Change. In M.H. Johnson (ed.), *Brain Development and Cognition: A Reader*. Oxford: Blackwell, 623–42.
- Bates, E., Thal, D. and Marchman, V. 1989: Symbols and Syntax: a Darwinian Approach to Language Development. Technical Report No. 89-04, Project in Cognitive and Neural Development, Neurodevelopmental Research Center, University of California at San Diego.
- Berthoud-Papandropoulou, I. 1978: An Experimental Study of Children's Ideas about Language. In A. Sinclair, R.J. Jarvella and W.J.M. Levelt (eds.), *The Child's Conception of Language*. Berlin: Springer-Verlag.
- Bever, T. 1992: The Logical and Extrinsic Sources of Modularity. Manuscript. University of Rochester.
- Bialystok, E. 1986a: Factors in the Growth of Linguistic Awareness. *Child Development*, 86, 498–510.
- Bialystok, E. 1986b: Children's Concept of Word. *Journal of Psycholinguistic Research*, 15, 13–32.
- Braine, M.D.S. 1987: What is Learned in Acquiring Word Classes? A Step towards an Acquisition Theory. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition*. Hillsdale, N.J.: Erlbaum.
- Brown, A.L. 1990: Domain-specific Principles Affect Learning and Transfer in Children. *Cognitive Science*, 14, 107–33.
- Brown, A.L., Bransford, J.D., Ferrara, R.A. and Campione, J.C. 1983: Learning, Remembering and Understanding. In P.H. Mussen (ed.), *Handbook of Child Psychology Volume 3: Cognitive Development*. New York: Wiley, 77–166.
- Chalmers, D. 1990: Syntactic Transformations on Distributed Representation. *Connection Science*, 2, 33–63.
- Clark, A. 1989: *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA.: MIT Press.
- Clark, A. 1990a: Connectionism, Competence and Explanation. *British Journal for the Philosophy of Science*, 41, 195–222.
- Clark, A. 1990b: Belief, Opinion and the Conscious Rule Interpreter. *Philosophical Psychology*, 3, 139–54.
- Clark, A. 1993. *Associative Engines: Connectionism, Concepts and Representational Change*. Cambridge, MA.: MIT Press.
- Clark, E.V. 1978: Awareness of Language: Some Evidence from What Children Say and Do. In A. Sinclair, R.J. Jarvella and W. Levelt (eds.), *The Child's Conception of Language*. Berlin: Springer-Verlag.
- Dennett, D. 1987: *The Intentional Stance*. Cambridge, MA.: MIT Press.
- Dennett, D. 1988: Science, Philosophy and Interpretation. *Behavioural and Brain Sciences*, 11, 535–46.
- Dennett, D. 1991: *Explaining Consciousness*. Boston: Little, Brown and Company.
- Elman, J.L. 1990: Finding Structure in Time. *Cognitive Science*, 14, 179–211.
- Elman, J.L. 1991: Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, 7, 195–225.

- Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.
- Finch, S. and Chater, N. 1991: A Hybrid Approach to the Automatic Learning of Linguistic Categories. *Artificial Intelligence and the Simulation of Behaviour CAISB7 Quarterly*, 78, 16-24.
- Fodor, J. and Pylyshyn, Z. 1988: Connectionism and Cognitive Architecture. *Cognition*, 28, 3-17.
- Freeman, N.H. 1980: *Strategies of Representation in Young Children: Analysis of Spatial Skills and Drawing Processes*. London: Academic Press.
- Freeman, N.H. 1987: Current Problems in the Development of Representational Picture Production. *Archives de Psychologie*, 55, 127-52.
- Hinton, G. 1988: Representing Part-whole Hierarchies in Connectionist Networks. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, Montreal, 48-54.
- Johnson-Laird, P. 1988: *The Computer and the Mind*. London: Fontana.
- Karmiloff-Smith, A. 1979a: *A Functional Approach to Child Language*. Cambridge: Cambridge University Press.
- Karmiloff-Smith, A. 1979b: Micro- and Macro-developmental Changes in Language Acquisition and Other Representational Systems. *Cognitive Science*, 3, 81-118.
- Karmiloff-Smith, A. 1984: Children's Problem Solving. In M.E. Lamb, A.L. Brown and B. Rogoff (eds.), *Advances in Developmental Psychology, Volume III*. Hillsdale, NJ.: Erlbaum.
- Karmiloff-Smith, A. 1985: Language and Cognitive Processes from a Developmental Perspective. *Language and Cognitive Processes*, 1, 61-85.
- Karmiloff-Smith, A. 1986: From Metaprocesses to Conscious Access: Evidence from Children's Metalinguistic and Repair Data. *Cognition*, 23, 95-147.
- Karmiloff-Smith, A. 1987: A Developmental Perspective on Human Consciousness. Invited Address, British Psychological Society Annual Conference, Sussex, April 1987.
- Karmiloff-Smith, A. 1990: Constraints on Representational Change: Evidence from Children's Drawing. *Cognition*, 34, 57-83.
- Karmiloff-Smith, A. 1992a: *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA.: MIT Press.
- Karmiloff-Smith, A. 1992b: Nature, Nurture and PDP: Preposterous Developmental Postulates? *Connection Science*, 4, 253-70.
- Karmiloff-Smith, A., Grant, J., Jones, M-C. and Cuckle, P. 1993: 'Rethinking Metalinguistic Awareness: Representing and Accessing Knowledge about What Counts as a Word'. *Discourse Processes*, 16, 565-89.
- Karmiloff-Smith, A., Johnson, H., Grant, J., Jones, M-C., Karmiloff, Y-N., Bartrip, J. and Cuckle, C. In press. From Sentential to Discourse Functions: Detection and Explanation of Speech Repairs by Children and Adults. *Discourse Processes*.
- Klahr, D., Langley, P., and Neches, R. (eds.) 1987: *Production System Models of Learning and Development*. Cambridge, MA.: MIT Press.
- Lachter, J. and Bever, T.G. 1988: The Relation between Linguistic Structure and Associative Theories of Language Learning: A Constructive Critique of Some Connectionist Learning Models. *Cognition*, 28, 195-247.
- Lehnert, W.G. 1987a: Case-based Problem Solving with a Large Knowledge Base of Learned Cases. *Proceedings of the Sixth National Conference on Artificial Intelligence*. Seattle, WA.
- Lehnert, W.G. 1987b: Word Pronunciation as a Problem in Case-based Reason-

- ing. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*. Seattle, WA.
- Maratsos, M. and Chalkley, M.A. 1980: The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories. In K. Nelson (ed.), *Children's Language 2*. New York: Gardner Press.
- McClelland, J.L. 1989: Parallel Distributed Processing: Implications for Cognition and Development. In R.G.M. Morris (ed.), *Parallel Distributed Processing: Implications for Psychology and Neurobiology*. New York: Oxford University Press.
- McClelland, J.L. and Jenkins, E. 1990: Nature, Nurture and Connectionism: Implications for Connectionist Models for Cognitive Development. In K. van Lehn (ed.), *Architectures for Intelligence*. Hillsdale, NJ.: Erlbaum.
- McClelland, J.L., Rumelhart, D.E. and the PDP Research Group 1986: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume 2: Psychological and Biological Models*. Cambridge, MA.: MIT Press.
- Mozer, M. and Smolensky, P. 1989: Using Relevance to Reduce Network Size Automatically. *Connection Science*, 1, 3-17.
- Newell, A. 1980: Physical Symbol Systems. *Cognitive Science*, 4, 135-83.
- Newell, A. 1990: *Unified Theories of Cognition*. Cambridge, MA.: Harvard University Press.
- Norman, D. 1986: Reflections on Cognition and Parallel Distributed Processing. In J.L. McClelland, D.E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume 2: Psychological and Biological Models*. Cambridge, MA.: MIT Press, 531-46.
- Pinker, S. and Prince, A. 1988: On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition. *Cognition*, 28, 73-193.
- Pinker, S. and Prince, A. Forthcoming: Rules and Connections in Human Language. *Trends in the Neurosciences*.
- Pollack, J. 1990: Recursive Distributed Representations. *Artificial Intelligence*, 46, 77-105.
- Plunkett, K. and Marchman, V. 1991: U-Shaped Learning and Child Language Acquisition. *Cognition*, 38, 43-102.
- Rumelhart, D.E. and McClelland, J.L. 1986: On Learning the Past Tenses of English Verbs. In J.L. McClelland, D.E. Rumelhart, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*. Cambridge, MA.: MIT Press, 216-71.
- Schultz, T.R. 1991: Simulating Stages of Human Cognitive Development with Connectionist Models. In L. Birnbaum and G. Collins (eds.), *Machine Learning: Proceedings of the Eighth International Workshop*. San Mateo, CA: Morgan Kaufmann, 105-109.
- Sejnowski, T. and Rosenberg, C. 1986: *NETtalk: a Parallel Network that Learns to Read Aloud*. Johns Hopkins University Electrical Engineering and Computer Science Technical Report, JHU/EEC-86/01.
- Servan-Schreiber, D., Cleeremans, A. and McClelland, J.L. 1988: Encoding Sequential Structure in Simple Recurrent Networks. CMU Technical Report CMU-CS-88-183. Computer Science Department, Carnegie-Mellon University.
- Shallice, T. 1988: *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

- Smolensky, P. 1987: Connectionist AI and the Brain. *Artificial Intelligence Review*, 1, 95-109.
- Smolensky, P. 1988: On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences*, 11, 1-74.
- Strauss, S. (ed.) 1982: *U-shaped Behavioral Growth*. New York: Academic Press.
- Thornton, C. 1991: Stirring Shakes: Introduction to Special Issue on Hybrid Models. *AISB Quarterly*, 78, 7.
- Touretsky, D. 1989: BoltzCONS: Dynamic Symbol Structures in a Connectionist Network. Carnegie-Mellon Computer Science Research Paper, CMU/CS/89/182.
- Touretsky, D. and Hinton, G. 1986: Symbols among the Neurons: Details of a Connectionist Inference Architecture. *Proceedings of 9th IJCAI*, Los Angeles, CA, 236-43.
- Tunmer, W.E., Bowey, J.A. and Grieve, R. 1983: The Development of Young Children's Awareness of the Word as a Unit of Spoken Language. *Journal of Psycholinguistic Research*, 12, 567-94.
- Tyler, L.K. 1988: Spoken Language Comprehension in a Fluent Aphasic Patient. *Cognitive Neuropsychology*, 5, 375-400.
- Van Sommers, P. 1984: *Drawing and Cognition: Descriptive and Experimental Studies of Graphic Production Processes*. Cambridge: Cambridge University Press.