Cambridge Handbook of Computational Cognitive Modeling (Chapter 15)

# Computational Models of Implicit learning

**Axel Cleeremans**

Université Libre de Bruxelles (U.L.B.)
Brussels, Belgium

**Zoltán Dienes**

University of Sussex
Brighton, U.K.

## 1. Introduction

Implicit learning — broadly construed, learning without awareness — is a complex, multifaceted phenomenon that defies easy definition. Frensch (1998) listed as many as eleven definitions in an overview — a diversity that is undoubtedly symptomatic of the conceptual and methodological challenges that continue to pervade the field 40 years after the term first appeared in the literature (Reber, 1967). According to Berry and Dienes (1993), learning is implicit when we acquire new information without intending to do so, and in such a way that the resulting knowledge is difficult to express. In this, implicit learning thus contrasts strongly with explicit learning (e.g., as when learning how to solve a problem or learning a concept), which is typically hypothesis-driven and fully conscious. Implicit learning is the process through which we become sensitive to certain regularities in the environment (1) without trying to learn regularities (2) without knowing that one is learning regularities, and (3) in such a way that the resulting knowledge is unconscious .

Over the last twenty years or so, the field of implicit learning has come to embody ongoing questioning about three fundamental issues in the cognitive sciences, namely (1) consciousness (how we should conceptualize and measure the relationships between conscious and unconscious cognition); (2) mental representation (in particular the

complex issue of abstraction); and (3) modularity and the architecture of the cognitive system (whether one should think of implicit and explicit learning as being subtended by separable systems of the brain or not). Computational modeling plays a central role in addressing these issues.

## 2. Implicit cognition: The phenomena

Everyday experience suggests that implicit learning is a ubiquitous phenomenon. For instance, we often seem to know more than we can tell. Riding a bicycle, using chopsticks or driving a car all involve mastering complex sets of motor skills that we find very difficult to describe verbally. These dissociations between our ability to report on cognitive processes and the behaviors that involve these processes are not limited to action but also extend to high-level cognition. Most native speakers of a language are unable to articulate the grammatical rules they nevertheless follow when uttering expressions of the language. Likewise, expertise in domains such as medical diagnosis or chess, as well as social or aesthetic judgments, all involve intuitive knowledge that one seems to have little introspective access to.

We also often seem to tell more than we can know. In a classic article, social psychologists Nisbett and Wilson (1977) reported on many experimental demonstrations that verbal reports on our own behavior often reflect reconstructive and interpretative processes rather than genuine introspection. While it is often agreed that cognitive processes are not in and of themselves open to any sort of introspection, Nisbett and

Wilson (1977) further claimed that we can sometimes be "(a) unaware of the existence of a stimulus that importantly influenced a response, (b) unaware of the existence of the response, and (c) unaware that the stimulus has affected the response". (p. 231).

Demonstrations of dissociations between subjective experience and various cognitive processes have now been reported in many domains of cognitive science. For instance, dissociations have been reported between conscious awareness and memory. Memory for previous events can be expressed explicitly, as a conscious recollection, or implicitly, as automatic, unconscious influences on behavior. Numerous studies have demonstrated dissociations between implicit and explicit memory, both in normal participants (see Schacter, 1987) as well in special populations. Amnesic patients, for instance, who exhibit severe or total loss in their ability to explicitly recall previous experiences (conscious recollection) nevertheless retain the ability to learn novel procedural skills or to exhibit sensitivity to past experiences of which they are not conscious.

Findings of "learning without awareness" have also been reported with normal subjects (Cleeremans, Destrebecqz, & Boyer, 1998). It is Arthur Reber, in a classic series of studies conducted in 1965 (see Reber, 1967), who first coined the term "mplicit learning" (though the phenomenon as such was discussed before Reber, for example, in Clark Hull's Ph.D. dissertation, published in 1920). Implicit learning contrasts with implicit memory in that implicit learning focuses on generalization to new stimuli rather

than sensitivity to processing the same stimulus again as such. Implicit learning also contrasts with subliminal perception in that it can involve consciously perceived stimuli.

Implicit learning research has essentially been focused on three experimental paradigms: Artificial Grammar Learning (henceforth, AGL), dynamic system control, and Sequence Learning (henceforth, SL). Additional paradigms that will not be discussed further include probability learning (Millward & Reber, 1968), hidden covariation detection, (Lewicki, 1986), acquisition of invariant characteristics (Lewicki, Hill, & Czyzewska, 1992), or visual search in complex stimulus environments (Chun & Jiang, 1999).

In Reber's seminal study of AGL (Reber, 1967), subjects were asked to memorize meaningless letter strings generated by a simple set of rules embodied in a finite-state grammar (Figure 1). After this memorization phase, subjects were told that the strings followed the rules of a grammar, and were asked to classify novel strings as grammatical or not. In this experiment and in many subsequent replications, subjects were able to perform this classification task better than chance despite remaining unable to describe the rules of the grammar in verbal reports. This dissociation between classification performance and verbal report is the finding that prompted Reber to describe learning as implicit, for subjects appeared sensitive to and could apply knowledge that they remained unable to describe and had had no intention to learn.

INSERT FIGURE 1 HERE

In a series of studies that attracted renewed interest in implicit learning, Berry and Broadbent (1984; 1988) showed that success in learning how to control a simulated system (e.g., a "sugar factory") so as to make it reach certain goal states was independent from ability to answer questions about the principles governing subject's inputs and the system's output: Practice selectively influenced ability to control the system, whereas verbal explanations about how the system works selectively influenced ability to answer questions.

Today, another paradigm — Sequence Learning — has become dominant in the study of implicit learning. In SL situations (Clegg, DiGirolamo, & Keele, 1998), participants are asked to react to each element of a sequentially structured visual sequence of events in the context of a Serial Reaction Time (SRT) task. On each trial, subjects see a stimulus that appears at one of several locations on a computer screen and are asked to press as fast and as accurately as possible on the key corresponding to its current location. Nissen and Bullemer (1987) first demonstrated that subjects progressively learned about the sequential structure of a repeating series of stimuli in spite of showing little evidence of being aware that the material was structured so. To establish that RT savings reflect sequence knowledge rather than mere familiarization with the task, a different sequence is typically presented during an unannounced transfer block, expected to elicit slower

reaction times to the extent that people use their knowledge of the sequence so as to anticipate the location of the next event. Cleeremans and McClelland (1991) used a different design in which the stimulus's location was probabilistically determined based on a finite-state grammar similar to that shown in Figure 1, and in which non-grammatical stimuli were randomly interspersed with those produced by the grammar. Numerous subsequent studies have indicated that subjects can learn about complex sequential relationships despite remaining unable to fully deploy this knowledge in corresponding direct tasks.

Most of the modeling work has focused on the AGL and the SL tasks and this chapter will therefore be focused on these paradigms (see Dienes & Fahey, 1995; Gibson, Fichman, & Plaut, 1997; Lebiere, Wallach, & Taatgen, 1998; Sun, 2002, for simulations of process control tasks). Both the AGL and SL tasks involve learning sequential dependencies and so involve similar computational problems. To put the computational modeling work in perspective and to highlight the challenging methodological and conceptual issues that characterize the domain, however, the next section is dedicated to discussing how to explore implicit learning empirically.

## 3. Demonstrating that implicit learning is implicit

The findings briefly reviewed above all suggest that unconscious influences on behavior are pervasive. This raises the question of how to best characterize the relationships between conscious and unconscious processes, and in particular whether one should

consider that mental representations can be unconscious. Settling the conceptual question of what conscious awareness is would help settle the methodological question of how to measure it. But there is no general agreement concerning what it means for an agent to be conscious of some state of affairs. There is a sense in which any perception of an object involves one being conscious of it. Thus, if by looking, a person can discriminate the presence or absence of an object then they are, in that sense, conscious of it being there. This sense of being 'conscious of' methodologically leads one to using direct forced choice tests as measures of awareness. If a person can discriminate whether an object is moving up or down when forced to say 'up' or 'down' on each trial, then in the sense we are talking about, the person is conscious of the object's direction of movement. In a similar way, if a person can discriminate whether a set of stimuli shared common features, one should conclude that they are conscious of that regularity. In this sense, subjects in implicit learning experiments are conscious of many regularities (e.g., Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990). For example, in AGL, subjects can indicate relevant parts of strings of letters that make them grammatical or non-grammatical (Dulany et al., 1984) and they can say whether particular bigrams (sequences of two letters) are allowed by a grammar or not (Perruchet & Pacteau, 1990). In SL, subjects can recognize parts or all of the sequence as old or new (e.g., Shanks & Johnstone, 1999). Further, in this sense of being conscious of regularities, the process of becoming conscious of the regularities can be simulated by computational models that learn to make the same discriminations as people do, as will be described in Section 4.

However, the useful distinction between implicit and explicit knowledge may not hinge on whether or not one is conscious of a regularity. It may hinge on whether a person is conscious of the regularity with a conscious rather than unconscious mental state. For example, in the sense we have been using, a blindsight patient is conscious of whether an object is moving up or down because the patient can discriminate direction of motion. But the seeing by which the patient is conscious of the object is not conscious seeing: The blindsight patient is conscious of the object with an unconscious mental state; one could say that the patient is sensitive to the object.

As a matter of general terminology, some people reserve the phrase 'conscious of' to cases where one is conscious of something by a conscious mental state; others use the phrase more generally, as we do here. In any case, there is now the problem of determining in what a mental state's being conscious consists. The conceptual answer to this question suggests both the methodology for determining whether people have conscious or unconscious knowledge in an implicit learning experiment, and the sort of computational model needed for simulating conscious rather than unconscious knowledge. Three approaches to defining the conscious status of mental states will be considered.

One approach claims that a mental state's being conscious is its being inferentially promiscuous, globally accessible (Baars, 1988; Block, 1995), or part of a suitable global

pattern of activity (Tononi & Edelman, 1998). According to this approach, a person has conscious knowledge of a regularity if that knowledge can be expressed in different ways, for example in verbal report or in different structured tests (Lewicki, 1986; Reber, 1967). The knowledge in implicit learning experiments is typically difficult to express in verbal report; indeed, this is the original finding that prompted Reber to conclude his AGL paradigm elicited unconscious knowledge. Further, the knowledge generated in implicit learning experiments can often be expressed only in some structured tasks but not others. For example, Jiménez, Mendez and Cleeremans (1996) measured the expression of knowledge learned through an SRT task using both reaction time and the ability to subsequently generate the sequence. Through detailed correlational analyses, they were able to show there was knowledge that was only expressed through the reaction time responses, but not through the sequence generation measure. The knowledge was thus not globally available. This type of study is more concincing that those using free report, as free report is often taken after a delay, without all retrieval cues present, and gives the subject the option of not reporting some conscious knowledge (Dulany, 1968). Thus, knowledge may be globally available yet not elicited on a testthat is insensitive or not asking for the same knowledge (Shanks & St. John, 1994). These issues can be addressed, for example, by asking subjects topredict the next element under the same conditions that they reacted to it, as Jimenez et al did. Computational models based on defining a conscious mental state in terms of global access include those of Tononi (2005) and of Dehaene and collaborators (e.g., Dehaene, Sergent, & Changeux, 2003), but will not be discussed further here.

Another approach is to identify conscious knowledge with knowledge that can be used according to one's intentions (Jacoby, 1991). This is a restricted form of inferential promiscuity that Jacoby operationally defines by his Process Dissociation Procedure. In the process dissociation procedure, a subject is asked in two different conditions ('inclusion' and 'exclusion') to do opposite things with a piece of knowledge. If the knowledge can be used according to opposing intentions, the knowledge is taken to be conscious (and unconscious otherwise). For example, Destrebecqz and Cleeremans (2001) applied the process dissociation procedure to SL, asking trained participants to either generate a sequence that resembled the training sequence (inclusion) or a sequence that was as different as possible from the training sequence (exclusion). Results indicated that while subjects could include the sequence when instructed, under certain conditions, participants were unable to exclude familiar sequence fragments, thus suggesting that they had no control over the knowledge acquired during training. Subjects could use the knowledge according to the intention to include but not the intention to exclude. Use was thus not determined by intentions. Destrebecqz and Cleeremans concluded that this knowledge was best described as implicit, for its expression was not under conscious control. They also produced a computational model of performance in the process dissociation task, discussed below (see also Tunney & Shanks, 2003; Vokey & Higham, 2004).

A third approach is to identify conscious mental states with states one is conscious of (Rosenthal, 2006); that is, with higher order states (i.e., mental states about mental states). On this approach, one must know that one knows for knowledge to be conscious. This approach suggest the use of subjective measures of awareness, such as confidence ratings. For example, a person may say, for each discrimination they perform in an AGL task, whether they were just guessing or whether they knew the correct answer. Two common criteria based on the confidence responses are the guessing and zero correlation criteria. According to the guessing criterion, if people can discriminate above chance when they believe they are guessing, the knowledge is unconscious. According to the zero correlation criterion, if people cannot discriminate with their 'guess' and 'know' responses between when they did and did not know, the knowledge is unconscious. According to both criteria, the knowledge acquired in AGL and SL paradigms is partly unconscious. The problem for computer simulation is to determine how a network could come to represent its own states as internal states and specifically as knowledge states. The problem is not trivial and as yet not fully resolved (Cleeremans, 2005).

Despite the considerable methodological advances achieved over the past decade or so, assessing awareness in implicit learning and related fields remains particularly challenging. There is no conceptual consensus on what a mental states' being conscious consists in, and hence no methodological consensus for determining the conscious status of knowledge. While the central issue of the extent to which information processing can

occur in the absence of conscious awareness remains as controversial today as it was 40 years ago, the conceptual and methodological tools are certainly more refined today.

A further challenge is to determine how to best interpret dissociations between conscious and unconscious knowledge in terms of systems or processes. Dunn and Kirsner (1988) pointed out that even crossed double dissociations between two tasks do not necessarily indicate the involvement of separable, independent processes. Many authors have described non-modular architectures that can nevertheless produce double dissociations. Plaut (1995) explored these issues in the context of cognitive neuropsychology. In a compelling series of simulation studies, Plaut not only showed that lesioning a single connectionist network in various ways could account for the double dissociations between concrete and abstract word reading exhibited by deep dyslexic patients, but also that lesions in a single site produced both patterns of dissociations observed with patients. In other words, the observed dissociations can clearly not be attributed to architectural specialization, but can instead be a consequence of functional specialization (functional modularity) in the representational system of the network. These issues are also debated in the context of implicit learning research. Computational modeling plays a key part in resolving such issues, just as it has in other domains. The process of implementing core conceptual ideas concerning the nature of conscious versus unconscious states together with ideas concerning the nature of human learning, testing implementations against human data, revising core concepts, and so on, cyclically, will help the field get beyond simple dichotomies. The brain is both in a sense one system, yet

it is also inhomogeneous. The verbal question of how many learning systems there are is in danger of being vacuous. If God were to tell us how many learning systems there were with a single number (one? two? three?), we would have learnt nothing. What we really need to know are the principles by which a working computational model of human learning could be built. It is early days yet, and models of implicit learning have focused more on the mechanisms of learning rather than on the conscious vs unconscious distinction (but see Sun, 2002). Future developments are eagerly awaited here.

## 4. Computational Models of implicit learning

Computational modeling has played a central role in deconstructing early verbal theories of the nature of what is learnt in implicit learning paradigms (1) by offering "proof of existence" demonstrations that elementary, associative learning processes (as opposed to rule-based learning) are in fact often sufficient to account for the data, (2) by making it possible to cast specific predictions that can then be contrasted with those of competing models, and (3) by making it possible to explore how specific computational principles can offer novel, unitary accounts of the data.

Detailed computational models have now been proposed for all three main paradigms of implicit learning. Two families of models are currently most influential: Neural network models, and fragment-based, or "chunking" models. Both approaches find their roots in exemplar-based models (Estes, 1957; Hintzmann, 1986; Medin & Schaffer,

1978), which had already captured the central intuition that rule-following behaviour can emerge out of the processing of exemplars in a germane domain — categorization.

Neural network models typically consist of simple auto-associator models (Dienes, 1992) or of networks capable of processing sequences of events, such as the Simple Recurrent Network (henceforth, SRN) introduced by Elman (1990) and first applied to SL by Cleeremans and McClelland (1991).

Fragment-based, or "chunking" models (e.g., Perruchet and Vinter, 1998), in contrast, are variants of exemplar-based models that assume that learning results in the acquisition of memory traces such as whole exemplars or fragments thereof.

While no type of model can currently claim generality, both approaches share a number of central assumptions: (1) learning involves elementary association or recoding processes that are highly sensitive to the statistical features of the training set, (2) learning is viewed essentially as a mandatory by-product of ongoing processing, (3) learning is based on the processing of exemplars and produces distributed knowledge, and (4) learning is unsupervised and self-organizing.

More recently, hybrid models that specifically attempt to capture the relationships between symbolic and subsymbolic processes in learning have also been proposed. Sun (2002), for instance, has introduced models that specifically attempt to link the

subsymbolic, associative, statistics-based processes characteristic of implicit learning with the symbolic, declarative, rule-based processes characteristic of explicit learning.

These different models have been essentially directed at addressing the questions of (1) what can be learned implicitly, and of (2) what are the computational principles characteristic of the mechanisms involved in implicit learning. In discussing the models , a third, important question will also be considered: How does one determine whether a model provides a good explanation of human learning? This issue is particularly acute in the domain of implicit learning, for there are often competing and overlapping accounts of the data. As an example, consider what could be learned based on having memorized a few letter strings from a finite-state grammar (Figure 2). People could learn about the rules that govern string generation; they could memorize a few frequent fragments of the training strings; they could learn about the statistical features of the material (e.g., the probability that each letter follows others); or they could simply memorize entire strings. Each of these possibilities would result in better-than-chance performance in a subsequent task asking participants to make decisions concerning the grammaticality of novel strings, and it remains a significant methodological challenge to design experimental situations that make it possible to successfully discriminate between the different competing accounts. Computational modeling is of great help in this respect for it forces the modeler to be explicit about his theory, but modeling raises its own challenge when it comes to comparing different models with a joint set of empirical data.

INSERT FIGURE 2 HERE

Section 5 is dedicated to considering the extent to which demonstrated dissociations between conscious and unconscious knowledge in people should be interpreted as reflecting the involvement of separable learning systems. Here, the basic features of the connectionist, chunking, and hybrid approaches are examined in turn.

**Connectionist Models of implicit learning**

The first fully implemented connectionist models of implicit learning are found in the early efforts of Dienes (1992) and of Cleeremans and McClelland (1991). While authors such as Brooks (1978) and Berry and Broadbent (1984) had already suggested that performance in implicit learning tasks such as AGL or Process Control may be based on retrieving exemplar information stored in memory arrays (see the Chapter by Logan, this volume), such models have in general been more concerned with accounting for performance at retrieval rather than on accounting for learning itself. The connectionist approach (see the Chapter by Thomas and McClelland, this volume), by contrast, has been centrally concerned with the mechanisms involved during learning since its inception, and therefore constitutes an excellent candidate framework with which to think about the processes involved in implicit learning. Because long-term knowledge in connectionist networks accrues in connection weights as a mandatory consequence of information processing, connectionist models capture, without any further assumptions, two of the most important characteristics of implicit learning, namely (1) the fact that

learning is incidental and mandatory, and (2) the fact that the resulting knowledge is difficult to express. A typical connectionist network, indeed, does not have direct access to the knowledge stored in connection weights. Instead, this knowledge can only be expressed through the influence that it exerts on the model's representations, and such representations may or may not contain readily accessible information (i.e., information that can be retrieved with no or low computational cost, see (i.e., information that can be retrieved with low or no computational cost, see Kirsh, 1991).

An important distinction in this regard is the distinction between supervised and unsupervised learning. O'Reilly and Munakata (2000) have characterized this distinction as a contrast between model learning (Hebbian, unsupervised learning) and task learning (error-driven, supervised learning). Their analysis is framed in terms of the different computational objectives the two types of learning fulfill: Capturing the statistical structure of the environment so as to develop appropriate models of it on the one hand, and learning specific input-output mappings so as to solve specific problems (tasks) in accordance with one's goals on the other hand. While many connectionist models of implicit learning have used supervised learning procedures, often, such models can also be interpreted as involving unsupervised learning (e.g., auto-associator networks).

Turning now to specific connectionist models of implicit learning, we will consider first a simple autoassociator as applied to AGL; then the more powerful SRN, which has

been applied to both SL and AGL tasks; and finally the memory buffer model, which has also been applied to both SL and AGL tasks.

**The autoassociator network.** Dienes (1992) proposed that performance in an AGL task could be accounted for based on the idea that, over training, people incidentally accumulate knowledge concerning the structure of the exemplars of the domain, and subsequently use that knowledge to make decisions concerning the grammaticality of novel exemplars in the transfer task. Dienes compared several instantiations of this basic idea in auto-associator networks trained with either the Hebb Rule or the Delta Rule.

In auto-associator networks, the task of the model is simply to reproduce the input pattern on its output units. The first problem in constructing a neural network is to decide how to encode the input. Dienes' models had no "hidden" units and used simple localist representation on both their input and output units, that is, each unit in the network represented the occurrence of a particular letter at a particular position in a string, or the occurrence of a particular bigram. The second problem is to decide what pattern of connection to implement. Dienes had each unit connected to all other units. That is, the network attempted to predict each unit based on all other units in the network. Finally, one has to decide what learning rule to use. Dienes used either the Hebb rule or the delta rule. The learning rules were factorially crossed with different coding schemes.

The two learning rules produce different types of knowledge. The Hebb rule, that is, the notion that "units that fire together wire together", learns the association between two units independently of any association those units may have with other units. After Hebbian learning the weights are like first-order correlations. The delta-rule, by contrast, involves competition between units in making predictions, so the weights are like multiple regression coefficients. The consequence was that for bigram models, the delta rule network could perfectly reproduce the training strings used and also any new string that could be formed by adding or subtracting any training strings. That is, simple associative learning produced rule-like behaviour, that is, perfect reproduction of any linear combination of the training strings without that rule being explicitly represented anywhere in the network — definitely one of the most important insights gained through connectionist modeling in this context.

All networks could classify test strings as well as people could: That is, all networks tended to reproduce grammatical test strings more faithfully than non-grammatical test strings. This raises a methodological problem: Why should one model be preferred over another as an account of human implicit learning? This question will be considered in the context of examining the different models of implicit learning that have been developed.

A key aspect of this problem is that networks in general have free parameters – numbers, like the learning rate, that have to be assigned some value for the network to give simulated behaviour. The delta rule network for example requires a learning rate;

different learning rates lead to different behaviours. Dienes dealt with this problem by producing parameter-free predictions. With a sufficiently small learning rate and sufficiently many training epochs the delta rule converges in the limit to producing multiple regression coefficients. The Hebb rule was parameter free in any case because it is a one-shot learning rule. The parameter-free models were tested by determining how well they predicted the order of difficulty human subjects had with classifying the strings. The delta rule model could predict the order of difficulty better than the Hebb rule[1].

The delta-rule autoassociator models passed the tests they were subjected to, but they have a couple of serious weaknesses. First, those models entail that people can learn to predict a letter in one position by the letters in any other position, no matter how far away; distance is irrelevant. But this entailment is false: People find long-distance dependencies in AGL hard to learn (Mathews et al., 1989). Second, those models entail that the association between two letters in two positions should not generalise to knowing the association between those letters in different positions. This entailment is very

[1] Dienes (1992) also considered variants of exemplar-based models (Estes, 1957; Hintzmann, 1986; Medin & Schaffer, 1978). These will not be elaborated on further here, but such models all share the assumption that grammaticality decisions are taken based on an item's similarity with the stored exemplars, accumulated over training with the material. These models turned out not to be good at predicting the order of difficulty of the test items, given the coding assumptions used.

unlikely. Cleeremans and McClelland (1991) simulated implicit learning with a connectionist model that dealt with both these problems.

**The Simple Recurrent Network.** Cleeremans and McClelland (1991) simulated performance in the SRT task. The network, Elman (1990)'s Simple Recurrent Network (SRN, see Figure 3), is a three-layer backpropagation network that is trained to predict each element of a sequence presented on its input units (see also the Chapter by Thomas and McClelland, this volume). Thus, on each trial, element t of a sequence is presented to the network (by activating a single input unit), and the network has to predict element t+1 of the sequence by activating the corresponding output unit. To make this prediction task possible, the network is equipped with so-called context units, which, on each time step through the sequence, contain a copy of the network's pattern of activity over its hidden units. Over time, the network learns to use these representations of its own activity in such a way as to refine its ability to predict the successor of each sequence element. Detailed analyses of the network's performance in learning sequential material have shown that the SRN's responses come to approximate the conditional probability of occurrence of each element in the temporal context set by its predecessors (Cleeremans, Servan-Schreiber, & McClelland, 1989).

---

INSERT FIGURE 3 HERE

---

Servan-Schreiber, Cleeremans and McClelland (1991) have shown that learning progresses through three qualitatively different phases when the network is trained on material generated from a finite-state grammar such as the one illustrated in Figure 1.

During a first phase, the network tends to ignore the context information. This is a direct consequence of the fact that the patterns of activation on the hidden layer — and hence the context layer — are continuously changing from one epoch to the next as the weights from the input units (the letters) to the hidden layer are modified. Consequently, adjustments made to the weights from the context layer to the hidden layer are inconsistent from epoch to epoch and cancel each other. In contrast, the network is able to pick up the stable association between each *letter* and all its possible successors. In a second phase, patterns copied on the context layer are now represented by a unique code designating which letter preceded the current letter, and the network can exploit this stability of the context information to start distinguishing between different occurrences of the same letter — different arcs in the grammar. Finally, in a third phase, small differences in the context information that reflect the occurrence of previous elements can be used to differentiate position-dependent predictions resulting from length constraints.

The internal representations that result from such training can be surprisingly rich and structured. Cluster analysis of the patterns of activation obtained over the network's hidden units after training on material generated from the probabilistic finite-state grammar revealed that the internal representations learned by the network are organized

in clusters, each of which corresponds to a node of the finite-state grammar. This turns out to be the most efficient representation of the input material from the point of view of a system that continuously attempts to predict what the next element will be, since knowing at which node a given sequence fragment terminates provides the best possible information concerning its possible successors. Just as the simple autoassociator considered by Dienes (1992) in some sense acquired abstract knowledge, so did the SRN. Cleeremans (1993) suggested it is useful to think of abstractness as lying on a continuum, and that verbal disputes over whether implicit knowledge is or is not abstract may be ill formed. The knowledge acquired by the SRN, in any case, has a level of abstractness somewhere between that of rote learning exemplars and learning the finite-state grammar propositionally.

As a model of human performance in SRT tasks, the SRN model has been shown to account for about 80% of the variance in the reaction time data (Cleeremans & McClelland, 1991). To capture reaction time data, one simply assumes that the normalized activation of each output unit is inversely proportional to reaction time. This is obviously a crude simplification, made necessary by the fact that back-propagation is unable to capture the time-course of processing. Other connectionist models have been more successful in this respect, such as Dominey's "Temporal Recurrent Network" (Dominey, 1998).

In modeling people's behavior with the SRN, there are a number of free parameters, including the learning rate, number of hidden units, and momentum. There is no easy way of obtaining parameter-free predictions. This is a methodological issue that will be addressed shortly in terms of what it means for assessing the SRN as an account for human learning.

The SRN model has also been applied to AGL tasks. For instance, Boucher and Dienes (2003) contrasted the SRN with a fragment-based model. Similarly, Kinder and Shanks used the SRN to model AGL in considering the question of how many learning systems there are (see Section 5) .

Both AGL and the SL task require the subject to learn sequential dependencies so it is not surprising that the same model has been brought to bear on the two tasks. It is an interesting question to what extent learning principles are the same in different domains of implicit learning. There is one key difference between AGL and SL stimuli, however. In AGL, the whole string is typically presented at once; in SRT, there is only one element of the sequence presented at a time. In fact, in AGL, performance decreases when the string is presented sequentially rather than simultaneously (Boucher & Dienes, 2003), implying that some modification of either coding or learning is needed when modelling standard AGL with the SRN. This point has not yet been addressed.

Dienes, Altmann and Gao (1999) considered a simple adaptation of the SRN in order to model the phenomenon of transfer between domains. Significantly, Reber (1969) showed that people trained on a finite-state grammar with one set of letters can classify new strings using a different set of letters (but the same grammar). The problem for the standard SRN is that the knowledge embedded in the connection weights is linked to particular letters. If new input units were activated, no previous learning would be relevant. Indeed, Marcus (2001) has regarded the inability to generalize outside the training space to be a general problem for connectionist models. Dienes et al. (1999) solved this problem by introducing an extra encoding layer between the input units and the hidden units, as shown in Figure 4.

---

INSERT FIGURE 4 HERE

---

In the training phase the network adjusts weights between the "domain one" input units all the way up to the "domain one" output units. The weights from the encoding layer to the hidden units, and from the context units to the hidden units — called the "core weights" — encode structural properties of the stimuli not tied to any particular letter set. In testing, the "domain two" input units are activated, and activation flows through the core weights to the output units. The core weights are frozen and the network learns the weights from the core part of the network to the input and output units in the new domain. Thus, the network learns how to best map the new domain onto the structures already present in its core weights. In this way, the network can indeed generalize outside

of its training space, and show various detailed properties shown by people (including infants) in transfer between domains in AGL. While the freezing of the core weights is simplistic, it shows connectionist networks can generalize beyond their training space. The freezing idea is similar to that of a switching device that determines how and when neural networks interface with each other (Jacobs, Jordan, & Barto, 1991).

Dienes et al. (1999) showed that the augmented SRN could predict a number of characteristics of human performance to within the 95% confidence limits of the effects. Fitting any more accurately would be fitting noise. Still, we must confront the methodological problem that the model has many free parameters. The required qualitative behaviour of the model was not restricted to a small region of parameter space. Nonetheless, simply showing a model can fit some behaviour is a weak scientific test. In general, if a model could produce a wide range of behaviour when the parameters are chosen appropriately, in what sense can the model *explain* any specific behaviour? Compare the exhortations of Popper that a *theory that can explain everything explains nothing,* and likewise *the more a theory rules out, the more it explains.* The discussion of the memory buffer model will methodologically squarely face up to these exhortations.

**The Memory Buffer Model.** The SRN is just one way, albeit an elegant way, of instantiating a memory buffer. The context units allow the SRN to learn (fallibly) how far into the past it needs a memory in order to reduce error. The SRN can be contrasted with a fixed memory buffer model, similar to the SRN in operating characteristics, learning

rule, etc, except for how time is coded. The architecture of the memory buffer model is similar to the SRN except that it has no context units (see Figure 5).

---

INSERT FIGURE 5 HERE

---

Rather than storing information about the previous events in the recurrent context units, the input units of the memory buffer model not only encode the input presented at time t, but also at time t-1, t-2, and t-3. The size of the memory buffer is specified by the number of time steps that are encoded. Moreover, the number of time steps that have been encoded will determine definitively the length of the non-local dependency that can be learnt. The simplicity of this means of encoding time (i.e. unfolded in space) has often recommended itself to researchers (see Sejnowski & Rosenberg, 1987, who developed NETtalk). Cleeremans (1993) fit a buffer network, coding four time steps into the past, to the reactions times of people learning the SRT task. He found people became gradually sensitive in their reaction times to information contained up to four time steps into the past, and the buffer network could behave in a similar way. The SRN and the buffer model were about equally good in this respect. He found that where the buffer model (with a buffer of four time steps) and the SRN made different predictions, and where the data differed significantly in that respect, the buffer model performed better than the SRN. Specifically, both the buffer model and people could learn a certain probabilistic difference over random intervening material, whereas the SRN could not.

Human learning in general requires a buffer. Aspects of language and music that can be learned in the lab rely on non-local dependencies, i.e. dependencies which take the form of two dependent items that are separated by a varying number of embedded items. Several studies have shown that under certain circumstances people can learn non-local dependencies that go beyond the learning of adjacent regularities. Kuhn and Dienes (2005) investigated the implicit learning of music with the AGL paradigm. People heard 8-note tunes in which the first note predicted the fifth, the second the sixth, and so on. (In fact, to be precise, the last four notes were the musical inversion of the first four.) After sufficient exposure, people came to like melodies respecting these mapping rules rather than other melodies. Some of the test melodies respecting the mapping rules had repeated sequences of notes from the training strings (the fragment set) and others were made from new note bigrams (the abstract set). People liked both sets equally; they had learnt the long-distance dependencies and this requires people had a buffer. But what sort of buffer do the mechanisms that subtend implicit learning use?

Kuhn and Dienes (2007) investigated how the SRN and the buffer network would learn the material. They found they with suitable encoding and parameter values both networks could fit the subjects level of performance. Figure 6 shows the behaviour of the SRN and memory buffer model over a full range of parameter values on both fragment and abstract test sets, with one input unit coding each musical note. The square in the figure represents a standard error above and below the human performance means. The SRN was relatively more sensitive to adjacent associations than long distance ones; the

fixed buffer model was equally sensitive to each. As people with these musical stimuli found the abstract and fragment sets equally difficult, the characteristic behaviour of the memory buffer model was more like that people than the characteristic behaviour of the SRN. In fact significantly more memory buffer models fell in the box defining human behaviour than SRN models.

---

INSERT FIGURE 6 HERE

---

With neural network models, one always has to consider whether different methods of coding the input would change the behaviour of the model. With different, more musically relevant coding schemes, more SRN models fell in the box defining human behaviour. That is, the SRN could fit the data. But there were always significantly more memory buffer models in the box than SRN models. The methodological moral is that in order to explain human data, find out if the model's characteristic behaviour matches that of people. The point is thus not so much whether the model can "fit" the data; rather, it is whether the model can explain the data because its processing principles and hypotheses about encoding entail a characteristic behaviour that matches that of people.

In sum, at this point, there is no clear "victor neural network model" of implicit learning. Perhaps the memory buffer model, though used in only two studies in the implicit learning literature, has an edge in SL and AGL applied to music. Future work needs to explore its use for AGL generally and whether it can be extended in the manner

of Dienes et al (1999) to allow transfer to different domains. However, it may be that different domains are learnt in different ways. People do not implicitly learn long distance contingencies with strings of letters very easily at all. What we learn about letters in everyday life is which letters chunk together, not what long distance dependencies there may be.

**Fragment-based models of implicit learning**

While connectionist models of implicit learning have been highly successful, one might argue that they fail to capture the fact that people, particularly in the AGL paradigm, typically perform a memorization task and hence end up consciously memorizing fragments, or chunks of the material. There is ample evidence that this knowledge is available for verbal report (Reber & Lewis, 1977) and it is therefore but a short step to assume that this knowledge is what drives people's ability to classify novel strings above chance (Perruchet & Gallego, 1997). These ideas are nicely captured by models that assume that learning involves accumulating fragmentary knowledge of the training material, and that performance at test involves using this knowledge to decide on the grammaticality of each novel string, for instance, by comparing its overlap in terms of fragments. The first such model was proposed by Servan-Schreiber and Anderson (1990) in the context of AGL. The model was called "Competitive Chunking" (CC). The central idea, well-known in the memory literature (Miller, 1956) but also in other domains (Newell, 1990) is that learning involves chunking of information: Production rules are combined so as to form larger units that execute faster; complex percepts are formed by

combining elementary features in different ways; items are committed to memory by organizing information so as to make it possible to exploit the redundancy of the material. In an AGL task, people asked to memorize meaningless letter strings chunk the material in short fragments (e.g., bigrams and trigrams). The Competitive Chunking model assumes that processing a letter string (or any other combination of elements) proceeds by recursively combining fragments of it until a single chunk can be used to represent the entire string. Thus for instance, a string such as TTXVPS might first be analyzed as (TT)X(VPS), then as (TT(X))(VPS), and then finally as ((TT(X))(VPS)). At this point, the entire string is represented as a single unit in the model and is said to be maximally familiar. Chunk formation in the model is a competitive process in which different potential chunks compete with each other: Each chunk receives bottom-up "support" from its constituent chunks, and its activation decays over time (see the Appendix for technical details). Servan-Schreiber and Anderson (1990) showed that competitive chunking offered a good account of performance in AGL tasks. More recently, Perruchet and Vinter (1998) have elaborated on these ideas by introducing a chunking model dubbed PARSER, based on similar principles (see the Appendix for further details concerning PARSER and a comparison with CC). While the model has so far not been applied in detail to implicit learning data, it shows great promise in capturing the fact that naturally come to perceive AGL strings as composed of chunks that they can report (Perruchet & Pacton, 2006; E. Servan-Schreiber & Anderson, 1990).

Boucher and Dienes (2003) contrasted the competitive chunker with the SRN as models of AGL. At one level both models learn the sequential dependencies produced by frequent bigrams and other chunks in the training strings. But their principles through which they lear are very different. The SRN is based on error-correction. If the bigram 'BV' occurs frequently in training, the SRN learns to predict that whenever B occurs, then V is likely to happen next. If BV no longer occurs but BX does, the SRN unlearns the BV connection and now comes to predict X given a B has occurred. This is a form of "catastrophic interference" (McCloskey & Cohen, 1989) that some neural networks are subject to. On the other hand, once the competitive chunker has learnt BV, it can then learn that BX is a chunk without unlearning that BV is also a chunk.

Boucher and Dienes presented people with training stimuli in which one bigram, BV, occurred in the first half of training and another in the last half. In the conflict condition, the other bigram was BX. In a control group PX occurred in the last half instead of BX. The question is to what extent did people unlearn in the conflict condition. People were asked to endorse different bigrams at the end of the test phase. The SRN and competitive chunker models were trained and tested in the same way over a full range of parameter values. Figure 7 shows the relative tendency of the models over a full range of parameter values to endorse bigram BV and also shows the mean value for people with confidence intervals.

INSERT FIGURE 7 HERE

Note the SRN is spread out all over the space; the competitive chunker's behaviour is more compact. Importantly, while the SRN models that could "fit" the data, it was the characteristic behaviour of the competitive chunker that best matched people's behaviour.

To summarise the main points so far, connectionist modelling is an excellent way of exploring theories of implicit learning. The SRN offers an elegant account of the data, but people show less interference and more sensitivity to long distance dependencies. Chunking models can capture the former, but not the latter (the long-distance dependencies learnt by Kuhn and Dienes's subjects cannot be learnt by current chunking models). A memory buffer model can capture the latter point but not the former, as it depends on error correction. Thus, there remains a problem of getting one model to exhibit all characteristics of human implicit learning! Perhaps this state of affairs will act as a spur to people interested in computational modelling. Finally, simple but important point is worth stressing: In comparing models, do not merely attempt to fit the data. Instead, look at the characteristic performance of models.

**Hybrid Models of Implicit Learning**

While the connectionist and fragment-based models reviewed above have proven extremely successful in accounting for implicit learning data, none have successfully addressed the central issue of how implicit knowledge may turn into explicit knowledge.

This, however, is a central issue in the cognitive sciences (Smolensky, 1988). Clark and Karmiloff-Smith (1993) pointed out that connectionist networks (and, by extension, any association-based model) have no "… self-generated means of analyzing 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" (p. 504). It is therefore a genuine, singular challenge, as Pinker (1999) suggests, to figure out how to best combine symbolic and subsymbolic approaches to cognition. In this respect, there are essentially four possible points of view about this, humorously summarized (from the perspective of die-hard connectionists) by Clark and Karmiloff-Smith (1993, pp. 504-505):

(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)

Recently, several models of implicit learning have been specifically directed at addressing the synergy between implicit and explicit learning. This approach makes a lot

of sense, for participants, even when placed in experimental situations designed to minimize the possibility of their becoming aware of the relevant regularities, will always attempt to infer explicit, conscious rules based on their experience of the situation. Further, many often also turn out to know something that they can verbalize about the material. In other words, one cannot simply turn awareness off, and there are good reasons to believe that performance in typical implicit learning situations always involve a mixture of implicit and explicit learning. Sun and colleagues (1997; 2002; Sun, Slusarz, & Terry, 2005) has attempted to address this issue by proposing a hybrid model of implicit learning called CLARION. The model uses both bottom-up, neural-network-based learning mechanisms and top-down, rule-based learning mechanisms. The model is thus genuinely hybrid and that it assumes continuous interaction between two separable components: One that is essentially symbolic in its representations and learning mechanisms, and another that is clearly sub-symbolic. Sun has applied CLARION to SL and process control tasks, simulating for instance in great detail the data of Curran and Keele (1993), which interestingly contrasted the influence of different instructions manipulating orientation to learn (i.e., incidental vs. intentional) in the task and the resulting differing degree of awareness of the material. Sun was able to capture these differences by manipulating the extent to which CLARION's symbolic component is allowed to extract rules from its subsymbolic component.

In the same spirit, Lebiere and collaborators (Lebiere et al., 1998; Wallach & Lebiere, 2000) have proposed ACT-R (Anderson, 1993) models of performance in SL and in

process control tasks. Learning in ACT-R (see the Chapter by Taatgen and Anderson, this volume). assumes that information processing is driven by the interaction between declarative knowledge structures (e.g., chunks of the stimulus material) and procedural knowledge, which in ACT-R take the form of production rules that implement the system's goals. The basic goal, for instance, in an SL situation, is to encode each stimulus and to respond to it using through a specific key. This and other productions operate on the declarative chunks acquired over training by the model, retrieving previously encoded chunks whenever appropriate to anticipate the location of the next stimulus. In such a model, explicit knowledge thus consists of the learned chunks, and implicit knowledge consists in the association strength between different co-occuring chunks that the model learns automatically.

Despite the appeal of hybrid models in accounting for the complex interactions between implicit and explicit learning (Domangue, Mathews, Sun, Roussel, & Guidry, 2004), detailed assessment of how well they compare with fragment-based and connectionist models in accounting for the human data must await further research.

## 5. Theoretical and conceptual implications

In this section, three central issues are addressed: Whether performance in implicit learning situations result in abstract knowledge, whether the data and the modeling suggest the involvment of single or multiple systems; and finally whether modeling is relevant to addressing the conscious vs. unconscious nature of the acquired knowledge.

**Rules vs. Statistics**. As discussed above, early characterizations of implicit knowledge have tended to describe it as abstract, based essentially on findings that subjects exhibit better-than-chance transfer performance, as when asked to make grammaticality judgments on novel strings in the context of AGL situations (Reber, 1989). Likewise, it has often been assumed that the reaction time savings observed in SRT tasks reflect the acquisition of "deep" knowledge about the rules used to generate the stimulus material (Lewicki, Czyzewska, & Hoffman, 1987). These abstractionist accounts have generally left it unspecified what the exact form of the acquired knowledge may be, short of noting that it must somehow represent the structure of the stimuli and their relationships, and be independent of the surface features of the material. The latter claim was further substantiated by findings that AGL knowledge transfers to strings based on the same grammar but instantiated with a different letter set, or even across modalities, as when training involves letter strings but transfer involves tone sequences.

However, as overviewed above, there now is considerable evidence that non-abstractionist mechanisms are largely sufficient to account for the data. Brooks (1978) first suggested that subjects in AGL experiments were classifying novel strings based not on abstract knowledge of the rules, but simply based on the extent to which novel grammatical or ungrammatical strings are similar to whole exemplars memorized during training. Perruchet and Pacteau (1990) showed that the knowledge acquired in both AGL and SL tasks might consist of little more than explicitly memorized short fragments or

chunks of the training material such as bigrams or trigrams, or simple frequency counts— which are perhaps the simplest form of abstraction. Both learning and transfer performance can then be accounted for by the extent to which novel material contains memorized chunks, as pointed out by Redington and Chater (1996; 2002), who emphasized that rule-like behaviour does not necessarily entail rule-based representations, — a point also made clear by many of the computational models reviewed here, such as Dienes et al. (1999)'s augmented SRN.

Overall, while it is clear that the knowledge acquired in typical implicit learning situations need not be based on the unconscious acquisition of symbolic rules, significant areas of debate remain about the extent to which unitary, fragment-based or associative mechanisms are sufficient to account for sensitivity to both the general and specific features of the training material. Simulation models have generally been suggestive that such mechanisms can in fact sufficient to account simultaneously for both grammaticality and similarity effects, partly because some instantiations of these mechanisms produce knowledge that lies on a continuum of abstractness. They can produce sets of weights that specify very precise rule-like behaviour (Dienes, 1992), that form graded finite-state patterns (Cleeremans, 1993), and that learn the specific lags over which dependencies occur (Kuhn & Dienes, in press) (Boyer, Destrebecqz, & Cleeremans, 2005).

The fact that both rule-based and exemplar-based approaches produce identical predictions over a large range of data is a significant issue that Pacton et al. (2001)

attempted to address by examining the untaught (and hence, incidental) acquisition of orthographic regularities over five years in a school setting. One prediction that rule-based approaches make is that after sufficient training, any acquired rules should generalize perfectly. Any learning mechanism based on the operation of associative learning mechanisms, however, would predict that performance on novel material will always lag behind performance on familiar material (the transfer decrement) These conditions are impossible to obtain in the laboratory, which motivated Pacton et al.'s longitudinal study. They found that performance on novel material indeed tended to lag, by a constant amount, behind performance on familiar material, a result that reinforces the idea that what people learn when they learn incidentally is essentially associative, rule-like knowledge, rather than rule-based knowledge.

**Separable systems?** Dissociations between implicit and explicit learning or processing have often been interpreted as suggesting the existence of separable memory systems. For instance, Squire and collaborators have shown that AGL is largely preserved in amnesia (e.g., Knowlton et al., 1992), to the extent that amnesic patients perform at the same level as normal controls when asked to classify strings as grammatical or not, but are impaired when asked to discriminate between familiar and novel instances (or fragments) of the strings. These results suggest that the processes that subtend declarative and non-declarative memory depend on separable brain systems respectively dedicated to representing either information about the specific features of each encountered exemplar

on the one hand (the hippocampus and related structures), and information about the features shared by many exemplars on the other hand (the neocortex).

In this case also however, computational modeling often casts the empirical findings in a different light. For instance, Kinder and Shanks (2001) were able to simulate the observed dissociations by tuning a single parameter (the learning rate) in an SRN trained on the same material as used in the behavioral studies, and therefore concluded that a single-system account is in fact sufficient to account for the data. The finding arises from the fact that the classification task and the recognition task were based on different test stimuli. The classification test consisted of new grammatical and new ungrammatical strings. The recognition task consisted of old grammatical and new grammatical material. The discriminations turned out to be differentially sensitive to changes in learning rate.

Not all learning by people consists of gradual change in sensitivity to distributional statistics, however. People consider possibilities and test hypotheses. The models overviewed in this chapter only function to model reality as it actually is. In the terms of Perner (1991), the models constitute 'single updating models'. As new information comes in, the model updates itself in an attempt to match reality more closely. The weights try to match the statistical structure of the world and the input units the occurrent stimulus. People, however, can, in Perner's terms, consider multiple models of the world; the real and the possible or the counterfactual. Our ability to engage with multiple models underlies much of our explicit learning. Integrating implicit and explicit learning

processes in a single model certainly deserves more work, following the example of Sun (2002).

**Conscious vs. unconscious knowledge.** As discussed in the Introduction, there is no sense in which current computational models can say much about the distinction between conscious and unconscious knowledge as observed in implicit learning tasks or, for that matter, in any other task (but see Dehaene et al., 2003; Mathis & Mozer, 1996, for interesting attempts). Nevertheless, there have been a few attempts at capturing the functional consequences of the distinction in terms of performance on different tasks (e.g., Sun, 2002, as discussed above). For instance, the SRN model as it stands fails to distinguish between anticipation and prediction responses, yet this difference is at the heart of the difference between the (largely implicit) facilitation observed when processing a sequence in the context of the SRT task and the (largely explicit) performance of participants asked to produce the same or a different sequence in the subsequent generation task. Destrebecqz and Cleeremans (2003) sought to address this limitation of the SRN by combining it with an auto-associator, so as to reflect the fact that people's task during the SRT task merely consists of mapping the current stimulus onto the correct response, whereas in the generation task they are expected to predict the location of the next element. The model was successful in capturing human data obtained over a range of conditions that either facilitated or promoted the acquisition of conscious knowledge. Likewise, Destrebecqz (2004) was able to capture the effects of manipulating orientation to learn and information both in an SRT task and on the subsequent

generation task by pretraining an SRN to different degrees, thus reflecting the idea that differences in availability to consciousness in this task reflects differences in the strength of the stored representations.

## 6. Conclusions

Implicit learning has proven to be a rich domain not only for the exploration of the differences between information processing with and without consciousness, but also for the development of computational models of the mechanisms involved in elementary learning. Because implicit learning situations typically involve incidental instructions, the mechanisms of change in such situations necessarily involve unsupervised processes that characterize learning as a by-product of information processing rather than as hypothesis-driven. Because the resulting knowledge is typically difficult to express, the most successful models all share the characteristic that they only involve elementary, associative learning mechanisms that result in distributed knowledge.

Based on the principles of successful models of implicit learning, it is appealing to consider it as a complex form of priming whereby experience continuously shapes memory, and through which stored traces in turn continuously influence further processing. Implicit learning studies suggest that such priming is far more interesting than the mere reinstatement of specific past experiences: The processes that produce it lead to quasi-abstract knowledge structures that allow the interesting generalizations that are at the heart of implicit learning.

Finally, while both fragment-based and neural network models make it clear how sensitivity to the distributional properties of an ensemble of stimuli can emerge out of the processing of exemplars, they differ in whether they assume that the shared features of the training materials are represented as such or merely computed when needed. This locus of abstraction issue is a difficult one that is unlikely to be resolved by modeling alone. Overall thus, it appears that the knowledge acquired through implicit learning is best described as lying somewhere on a continuum between purely exemplar-based representations and more general, abstract representations — a characteristic that neural network models have been particularly apt at capturing. Further research is needed to develop unified models of implicit learning, and to gain insight into the computational principles that differentiate conscious from unconscious processing.

## Acknowledgements

A.C. is a Research Director with the Fund for Scientific Research (F.R.S.-F.N.R.S., Belgium). This work was supported by an institutional grant from the Université Libre de Bruxelles to A.C., by Concerted Research Action 06/11-342 titled "Culturally modified organisms: What it means to be human in the age of culture", financed by the Ministère de la Communauté Française – Direction Générale l'Enseignement non obligatoire et de la Recherche scientifique (Belgium), and by F.R.F.C. grant 2.4577.06.

## References

Anderson, J. R. (1993). *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology, 36A*, 209-231.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79*, 251-272.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove, UK: Lawrence Erlbaum Associates.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227-287.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science, 27*, 807-842.

Boyer, M., Destrebecqz, A., & Cleeremans, A. (2005). Processing abstract sequence structure: Learning without knowing, or knowing without learning? *Psychological Research, 69*, 383-398.

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and Concepts* (pp. 16-211). Mahwah, N.J.: Lawrence Erlbaum Associates.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10*, 360-365.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language, 8*, 487-519.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.

Cleeremans, A. (2005). Computational correlates of consciousness. In S. Laureys (Ed.), *Progress in Brain Research* (Vol. 150, pp. 81-98). Amsterdam: Elsevier.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2*, 406-416.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology : General, 120*, 235-253.

Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite state automata and simple recurrent networks. *Neural Computation, 1*, 372-381.

Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in Cognitive Sciences, 2*, 275-281.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology : Learning, Memory and Cognition, 19*, 189-202.

Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the U.S.A., 100*(14), 8520-8525.

Destrebecqz, A. (2004). The effect of explicit knowledge on sequence learning: A graded account. *Psychological Belgica, 44*(4), 217-248.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the Process Dissociation Procedure. *Psychonomic Bulletin & Review, 8*(2), 343-350.

Destrebecqz, A., & Cleeremans, A. (2003). Temporal effects in sequence learning. In L. Jiménez (Ed.), *Attention and Implicit Learning* (pp. 181-213). Amsterdam: John Benjamins.

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science, 16*, 41-79.

Dienes, Z., Altmann, G., & Gao, S.-J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science, 23*, 53-82.

Dienes, Z., & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 21*, 848-862.

Domangue, T., Mathews, R. C., Sun, R., Roussel, L. G., & Guidry, C. (2004). The effects of model-based and memory-based processing on speed and accuracy of grammar string generation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*(5), 1002-1011.

Dominey, P. F. (1998). Influences of temporal organization on sequence learning and transfer: Comments on Stadler (1995) and Curran and Keele (1993). *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 234-248.

Dulany, D. E. (1968). Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In T. Dixon & D. Horton (Eds.), *Verbal Behavior and Behavior Theory* (pp. 340-387). New York, NY: Prentice-Hall.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgement : How conscious and how abstract? *Journal of Experimental Psychology : General, 113*, 541-555.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental process: The principle of reversed association. *Psychological Review, 95*, 91-101.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Estes, W. K. (1957). Toward a statistical theory of learning. *Psychological Review, 57*, 94-107.

Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 47-104). Thousand Oaks, CA: Sage Publications.

Gibson, F., Fichman, M., & Plaut, D. C. (1997). Learning in dynamic decision task: Computational models and empirical evidence. *Organizational Behavior and Human Decision Processes, 71*, 1-35.

Hintzmann, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where of vision tasks. *Cognitive Science, 15*, 219-250.

Jacoby, L. L. (1991). A process dissociation framework : Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Jiménez, L., Mendez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(4), 948-969.

Kinder, A., & Shanks, D. R. (2001). Amnesia and the Declarative/Nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience, 13*(5), 648-669.

Kirsh, D. (1991). When is information explicitly represented? In P. P. Hanson (Ed.), *Information, Language, and Cognition*. New York, NY: Oxford University Press.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of non-local musical rules: Implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(6), 1417-1432.

Kuhn, G., & Dienes, Z. (in press). Learning non-local dependencies. *Cognition*.

Lebiere, C., Wallach, D. P., & Taatgen, N. A. (1998). Implicit and explicit learning in ACT-R. In F. Ritter & R. Young (Eds.), *Cognitive Modeling* (Vol. II, pp. 183-193). Nottingham: Notthingam University Press.

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology : Learning, Memory and Cognition, 12*, 135-146.

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology : Learning, Memory and Cognition, 13*, 523-530.

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *American Psychologist, 47*, 796-801.

Marcus, G. F. (2001). *The Algebraic Mind. Integrating Connectionism and Cognitive Science*. Cambridge, MA: MIT Press.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit process in learning from examples: A synergistic effet. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 1083-1100.

Mathis, W. D., & Mozer, M. C. (1996). Conscious and unconscious perception: A computational theory. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 324-328). Hillsdale, N.J.: Lawrence Erlbaum Associates.

McCloskey, M., & Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 24, pp. 109-164). New York, NY: Academic Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Miller, G. A. (1956). The magical number Seven, plus or minus two. *Psychological Review, 63*, 81-97.

Millward, R. B., & Reber, A. S. (1968). Event-recall in probability learning. *Journal of Verbal Learning and Verbal Behavior, 7*, 980-989.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge: Harvard University Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can do: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirement of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1-32.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130*(3), 401-426.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perruchet, P., & Gallego, G. (1997). A subjective unit formation account of implicit learning. In D. Berry (Ed.), *How implicit is implicit knowledge?* (pp. 124-161). Oxford: Oxford University Press.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning : Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology : General, 119*, 264-275.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: Two approaches, one phenomeon? *Trends in Cognitive Sciences, 10*, 233-238.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17*, 97-119.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*, 246-263.

Pinker, S. (1999). Out of the mind of babes. *Science, 283*, 40-41.

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropschology, 17*, 291-326.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 5*, 855-863.

Reber, A. S. (1976). Implicit learning of synthetic languages. *Journal of Experimental Psychology : Human Learning and Memory, 2*, 88-94.

Reber, A. S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition, 114*, 14-24.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General, 125*, 123-138.

Redington, M., & Chater, N. (2002). Knowledge representation and transfer in artificial grammar learning. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness*. Hove, UK: Psychology Press.

Rosenthal, D. (2006). *Consciousness and Mind*. Oxford, UK: Oxford University Press.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501-518.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems, 1*, 145-168.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J. L. (1991). Graded State Machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning, 7*, 161-193.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammar with competitive chunking. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 592-608.

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a serial reaction time task. *Journal of Experimental Psychology : Learning, Memory, & Cognition.*

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367-447.

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

Sun, R. (1997). Learning, action, and consciousness: A hybrid approach towards modeling consciousness. *Neural Networks, 10*(7), 1317-1331.

Sun, R. (2002). *Duality of the Mind*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the implicit and the explicit in skill learning: A dual-process approach. *Psycological Review, 112*(1), 159-192.

Tononi, G. (2005). Consciousness and the brain: Some theoretical considerations. In S. Laureys (Ed.), *Progress in Brain Research* (Vol. 150, pp. xx). Amsterdam: Elsevier.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science, 282*(5395), 1846-1851.

Tunney, R. J., & Shanks, D. R. (2003). Does opposition logic provide evidence for conscious and unconscious processes in artificial grammar learning? *Consciousness and Cognition, 12*, 201-218.

Vokey, J. R., & Higham, P. A. (2004). Opposition logic and neural network models in artificial grammar learning. *Consciousness and Cognition, 3*, 565-578.

Wallach, D. P., & Lebiere, C. (2000). Learning of event sequences: An architectural approach. In N. A. Taatgen (Ed.), *Proceedings of the Third International Conference on Cognitive Modeling* (pp. 271-279). Gröningen: Universal Press.

# Appendix

We present here the equations for the two main chunking models in the implicit learning literature, the Competitive Chunker (CC) of Servan-Scheiber and Anderson (1990) and the PARSER model of Perruchet and Vinter (1998).

**Competitive Chunker.** CC perceives a stimulus by successively chunking together the basic components of that stimulus until a single chunk represents it. So, using brackets to denote a chunk, the exemplar 'MTVR' might be perceived as first as 'MTVR' – i.e. as '(M)(T)(V)(R)', then '(MT)VR', then '(MT)(VR)' and finally '((MT)(VR))'. Once a stimulus is fully chunked it is said to be maximally familiar, or memorised.

Initially CC is given elementary chunks, e.g. letters. Each chunk has a strength. Strength is increased by one unit every time the chunk is used or recreated. However, strength decays with time. At any point in time, the strength of a chunk is the sum of its successive individually decaying strengthenings:

$$\text{strength} = \Sigma_i \, T_i^{-d} \qquad (1)$$

where $T_i$ is the time elapsed since the ith strengthening and *d* is the *decay parameter* ($0<d<1$).

Given 'MTVR' it will consider all possible combinations of two adjacent existing chunks as possible new chunks, i.e. 'MT', 'TV', and 'VR'. Each possibility has a support, given by the sum of the strengths of each of its subchunks. The probability that a new chunk will be formed is given by:

$$(1 - e^{-c*support})/(1 + e^{-c*support}) \quad (2)$$

where c is the *competition parameter*, $c > 0$. Only one new chunk is formed at a time. Thus, the three chunks 'MT', 'TV' and 'VR' will compete with each other to be created. If 'MT' is formed as a chunk, next time the stimulus is seen, possible new chunks are 'MTV', and 'VR', which will compete to be formed by the same process.

When a stimulus is presented, the mere existence of a chunk that matches part of the stimulus does not mean it will be retrieved. The probability of retrieving a chunk is given by equation (2), the same equation as for chunk creation. Thus it may be that two competing chunks are retrieved, e.g. both 'MTV' and 'VR'. In that case, the stronger chunk wins. The greater the value of c, the more likely it is that chunks will be retrieved, and hence the greater the probability of competition. After a first pass, another pass is made to see if the existing chunks can be perceived as higher-order chunks. At a certain point no further chunks are retrieved. At this stage, if the resulting percept is not one single chunk, a further chunk may be created, as described.

The familiarity of a stimulus is given by the number of active resulting from the perceptual process, e.g.:

$$\text{familiarity} = e^{1-n_{active}} \quad (3)$$

This familiarity value can then be used to classify strings as grammatical, old, etc.

**PARSER.** Like CC, PARSER begins with a set of primitives, e.g. letters. When presented with a string like 'MTVRXX' it randomly considers perceiving groups of 1, 2 or 3 primitives reading from left to right. (PARSER differs from CC in parsing from left to right: PARSER was originally used to model the perception of auditory strings and CC was developed to model visual strings.) For example, if it randomly produced '1, 3, 2' it would see the string as (M)(TVR) (XX). Because TVR and XX do not exist as units, they become new perceptual units and are assigned weights (like CC's strengths) (for example, all new units could be assigned weights of 1). 'M' already exists and its weight is incremented (by an amount *a*). At each time step all units are affected by forgetting and interference. Forgetting is simulated by decreasing all the units by a fixed value *f*. Interference is simulated by decreasing the weights of the units in which any of the letters involved in the currently processed unit are embedded (by an amount *i*). Once new units have been formed, they act in the cycle above just like primitive units. All units can contribute to perception so long as their weight exceeds a threshold (*t*). As for CC, the number of chunks a string is perceived as could be used to determine its familiarity.

**Comparison.** CC and PARSER both postulate that learning occurs by chunking in which (a) the use of a chunk increments its weight; and (b) each chunk decays in weight on each time step. They theoretically differ in that (c) PARSER, but not CC, has an interference process by which chunks that are not used but that contain an element that was used are decremented in weight. Because of (a) and (b), both models correctly predict that with the strengthening of common chunks and fading of infrequent ones, people will come to perceive stimuli as made of the commonly occurring chunks.

PARSER's interference parameter has two effects. One is that it tends to eliminate long items (long items are obviously very prone to interference, because many small items interfere with them). But perhaps more importantly, it makes PARSER sensitive to both forward transitional probabilities (the conditional probability of a second event given a first) and backward transitional probabilities (the conditional probability of a first event given a second). CC is mainly sensitive to the frequency of co-occurrence of two items next to each other rather than transitional probabilities. The SRN is sensitive to forward but not backward transitional probabilities. Perruchet and Peereman (2004) showed that in rating the goodness of non-words as being words, people were sensitive to both forward and backward transitional probabilities, consistent with PARSER but neither with the SRN nor with CC. Further, in many statistical learning situations, people are sensitive to transition probabilities (e.g., Aslin, Saffran, & Newport, 1998). Conversely, Boucher and Dienes (2003) found support for CC over the SRN in artificial grammar learning because people were mainly sensitive to co-occurrence frequency.

Thus, it is likely PARSER could fit the Boucher and Dienes data by letting the interference parameter go to 0, but that would be an ad hoc solution because PARSER's characteristic behavior is sensitivity to transition probabilities. Nonetheless PARSER provides a framework for future research to establish a meaningful way of indicating when its interference parameter should go to 0 and when it should not.

## Figure captions

**Figure 1:** A finite-state grammar (Reber, 1976) is a simple directed graph consisting of nodes connected by labeled arcs. Sequences of symbols can be generated by entering the grammar through a "begin" node, and by moving from node to node until an "End" node is reached. Each transition between a node and the next produces the label associated with the arc linking the two nodes. Concatenating the symbols together produces strings of symbols, in this case, letters of the alphabet. Finite-state grammars have been used both in the context of Sequence Learning studies and in the context of Artificial Grammar Learning studies.

**Figure 2:** A representation of different computational approaches to Artificial Grammar Learning (see text for details).

**Figure 3:** The Simple Recurrent Network (SRN) introduced by Elman (1990). The network takes the current element of a sequence as input, and is trained to predict the next element using back-propagation. Context units, which on time step contain a copy of the activation pattern that existed over the network's hidden units on the previous time step, enable previous information to influence current predictions.

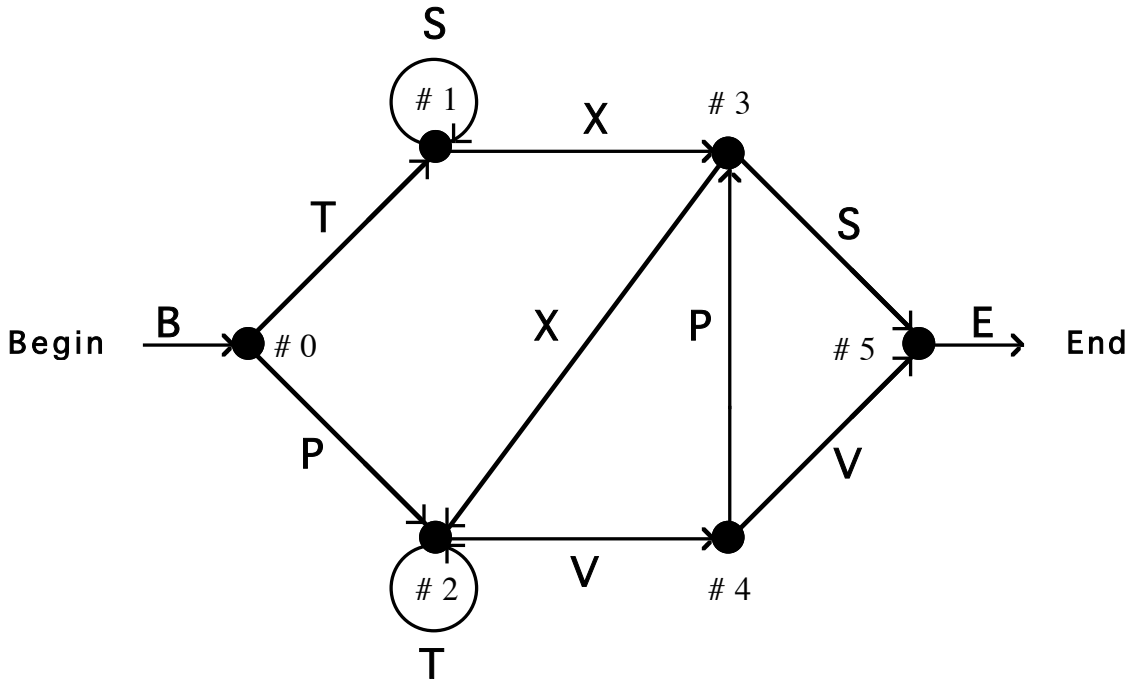
**Figure 4:** The model of Dienes, Altmann and Gao (1999). Transfer between domains is achieved by augmenting an SRN network with "mapping" weights that make it possible

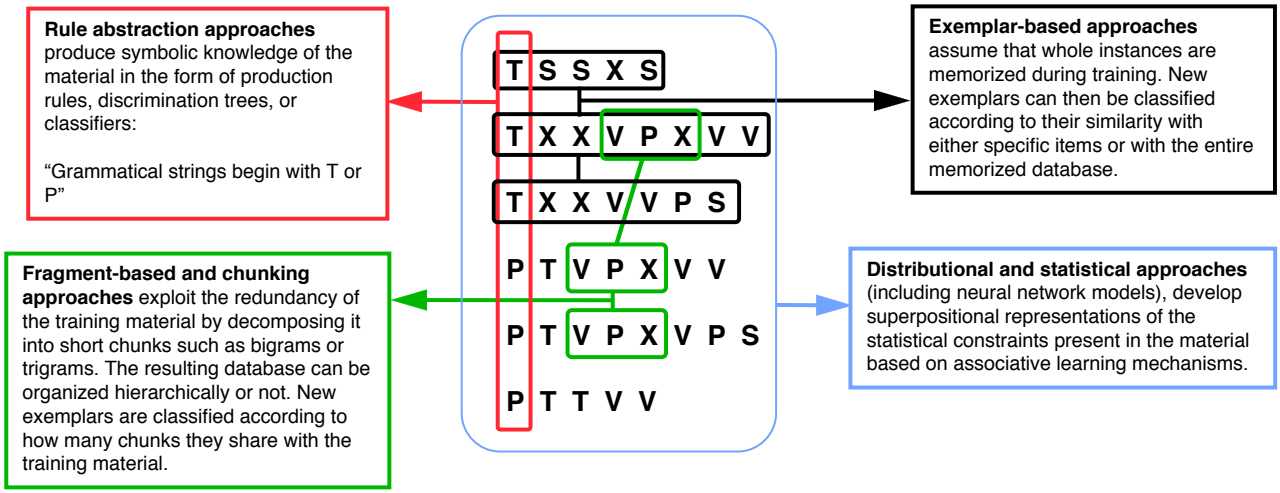
for the knowledge embedded in the "core" weights to be preserved and used for generalization when switching to a different set of stimuli.

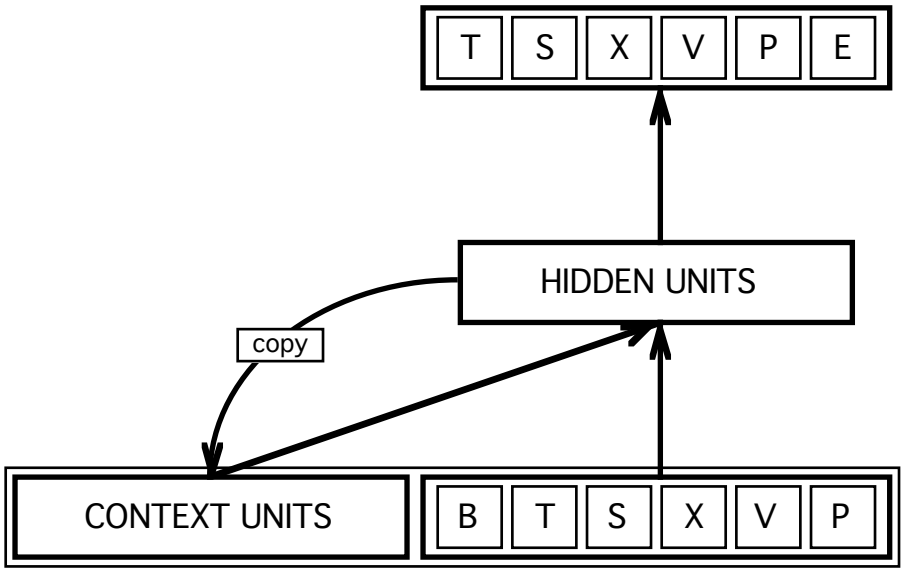
**Figure 5:** The buffer network. A fixed-width time window is implemented by input units dedicated for each time slot.

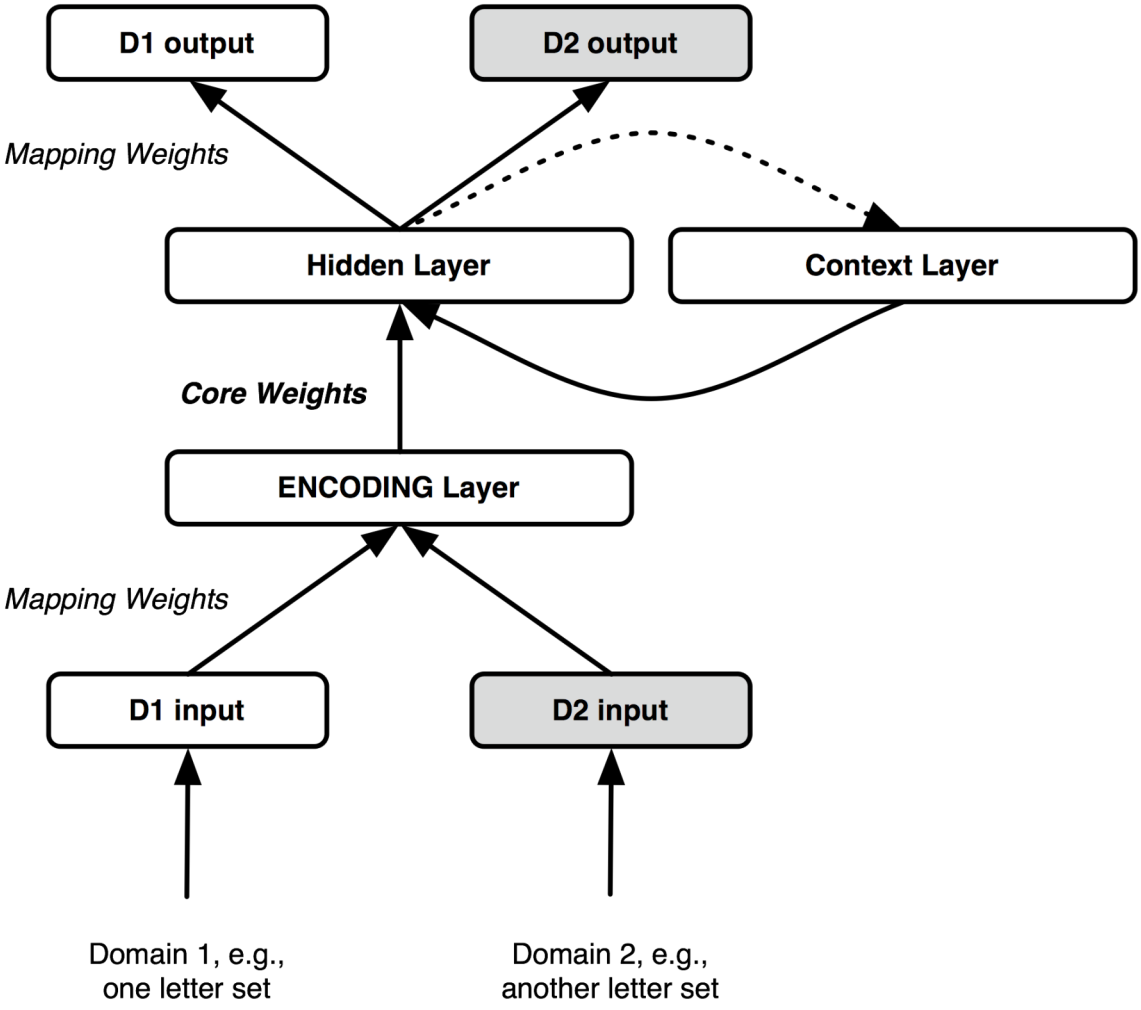
**Figure 6:** Performance of the memory buffer network and of the SRN on music stimuli over a full range of parameters. The box shows a standard error above and below human means. The buffer network is characteristically more like human behavior than the SRN is. See text for full explanation.

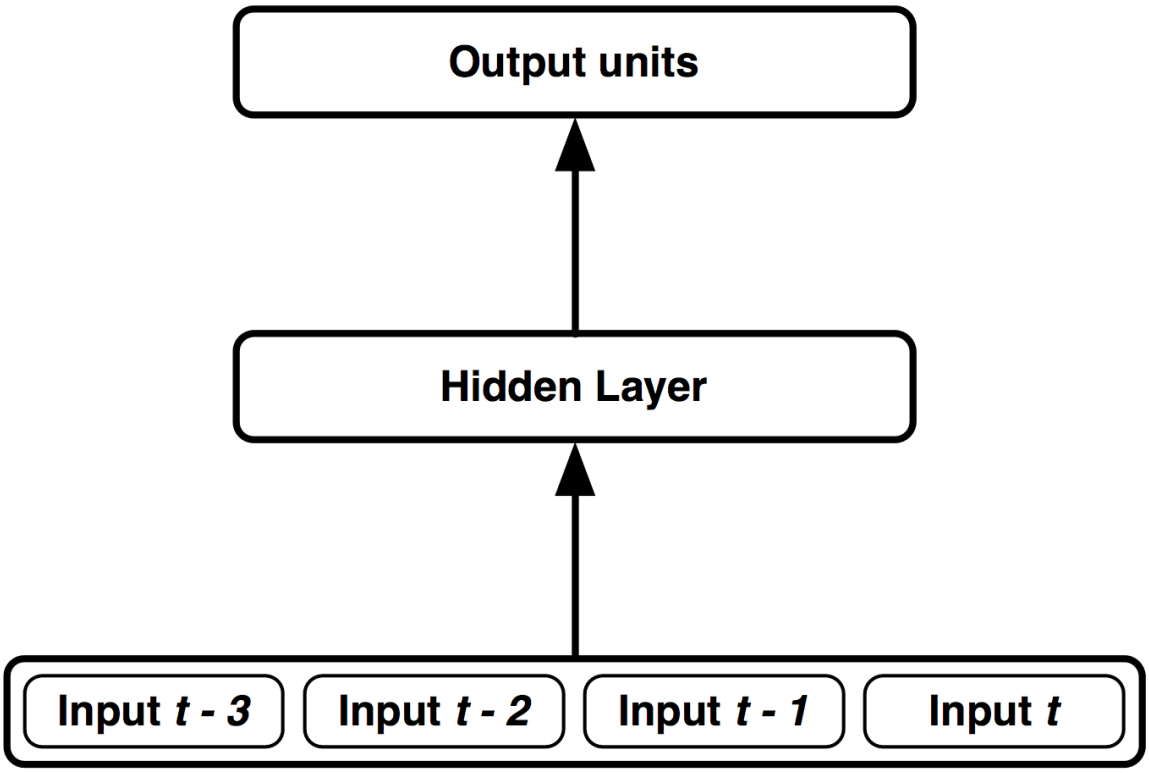
**Figure 7:** Performance of the competitive chunker and SRN models in dealing with prediction conflicts. The competitive chunker is resistant to conflict whereas the SRN shows a range of sensitivity to it. Humans are resistant, like the competitive chunker.See text for full explanation.

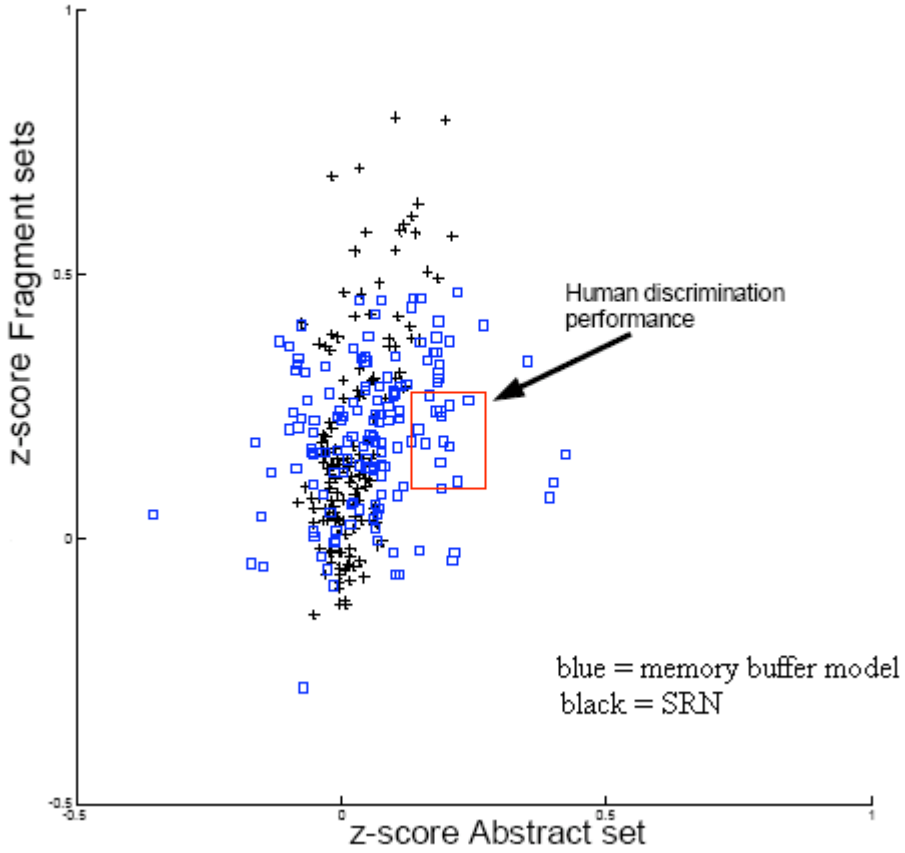

S
# 1
X
# 3
T
S
Begin
B
# 0
X
P
# 5
E
End
P
V
V
# 2
# 4
T

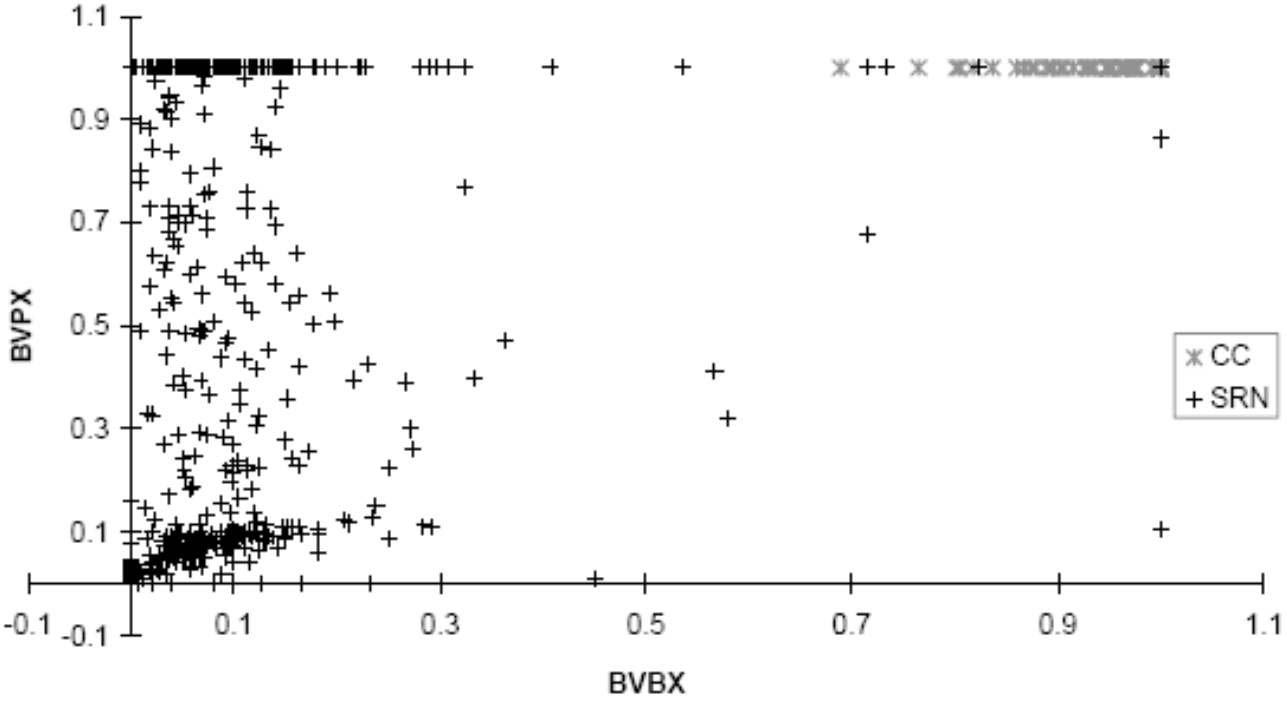

Rule abstraction approaches produce symbolic knowledge of the material in the form of production rules, discrimination trees, or classifiers:
"Grammatical strings begin with T or P"
Fragment-based and chunking approaches exploit the redundancy of the training material by decomposing it into short chunks such as bigrams or trigrams. The resulting database can be organized hierarchically or not. New exemplars are classified according to how many chunks they share with the training material.
T S S X S
T X X V P X V V
T X X V V P S
P T V P X V V
P T V P X V P S
P T T V V
Exemplar-based approaches assume that whole instances are memorized during training. New exemplars can then be classified according to their similarity with either specific items or with the entire memorized database.
Distributional and statistical approaches (including neural network models), develop superpositional representations of the statistical constraints present in the material based on associative learning mechanisms.

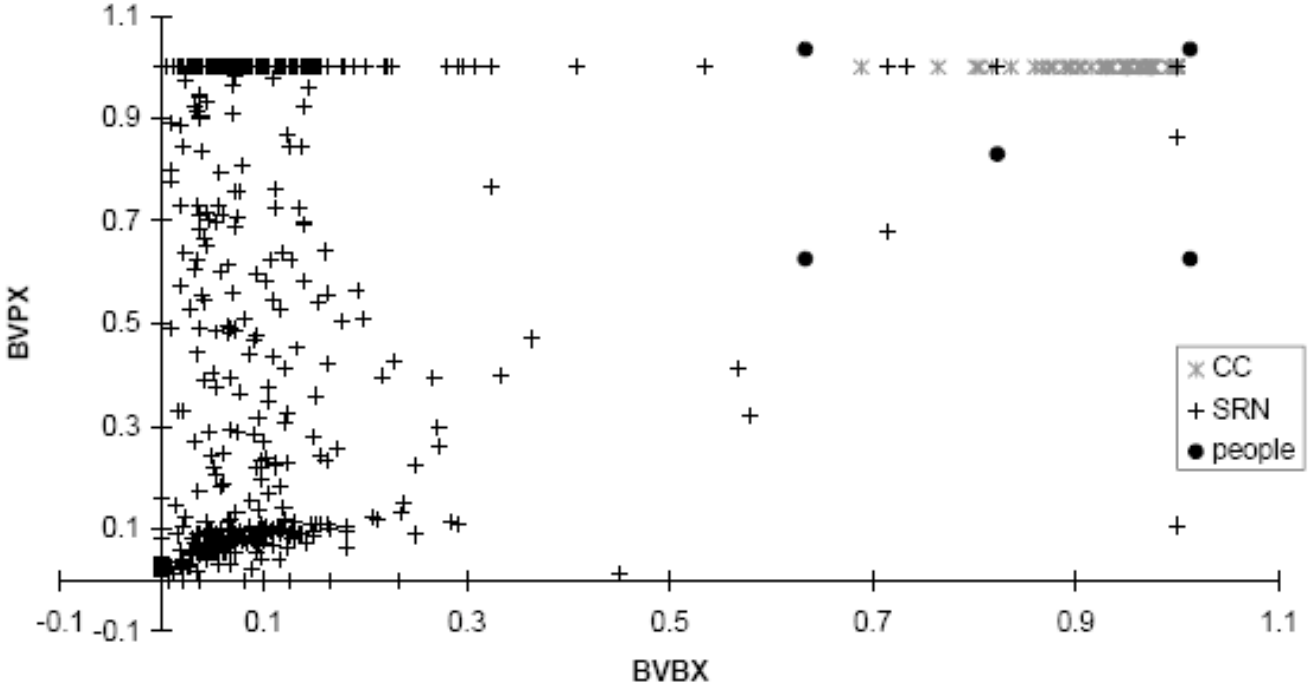

T
S
X
V
P
E
HIDDEN UNITS
copy
CONTEXT UNITS
B
T
S
X
V
P


D1 output
D2 output
Mapping Weights
Hidden Layer
Context Layer
Core Weights
ENCODING Layer
Mapping Weights
D1 input
D2 input
Domain 1, e.g.,
one letter set
Domain 2, e.g.,
another letter set


Output units
Hidden Layer
Input t - 3
Input t - 2
Input t - 1
Input t


z-score Fragment sets
z-score Abstract set
Human discrimination performance
blue = memory buffer model
black = SRN
1
0.5
0
-0.5
-0.5
0
0.5
1


1.1
0.9
0.7
0.5
0.3
0.1
-0.1
-0.1
0.1
0.3
0.5
0.7
0.9
1.1
BVPX
BVBX
CC
SRN


1.1
0.9
0.7
0.5
0.3
0.1
-0.1
-0.1
0.1
0.3
0.5
0.7
0.9
1.1
BVPX
BVBX
CC
SRN
people